

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2001, 15:36:17 ; Search time 49.95 Seconds

(Without alignments)
279.236 Million cell updates/sec

Title: US-09-543-371-10

Perfect score: 1340

Sequence: 1 GLKGRGDSGSPATWTTRGF.....KAGELEKITSRCVCMKKRH 244

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 segs, 57163235 residues

T number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0401.*
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1340	100.0	245	21	Y67942 Human type IV coll
2	1314	98.1	268	20	Y31893 Type IV collagen N
3	1314	98.1	268	21	Y97555 Human alpha3(IV)NC
4	1210.5	90.3	471	20	Y44171 Bovine type IV col
5	1210.5	90.3	471	21	Y56783 Bovine alpha3 type
6	1199.5	89.5	471	16	R79163 Partial sequence o
7	1192	89.0	218	20	Y44172 Human type IV coll
8	1192	89.0	218	21	Y56784 Human alpha3 type
9	1170	87.3	218	16	R79164 Partial sequence o
10	960	71.6	406	21	B58169 Lung cancer associ
11	943.5	70.4	772	13	R23873 Human alpha 5 (IV)

12	943.5	70.4	772	18	W09643	Human type IV coll
13	939	70.1	229	21	Y67943	Human type IV coll
14	939	70.1	260	20	Y31891	Type IV collagen N
15	939	70.1	260	21	Y97553	Human alpha1(IV)NC
16	938.5	70.0	264	20	Y31895	Type IV collagen N
17	938.5	70.0	264	21	Y97557	Human alpha5(IV)NC
18	937.5	70.0	309	21	B54044	Human pancreatic c
19	923	68.9	229	10	P93524	Complete sequence
20	880	65.7	211	21	Y95918	Human Goodpasture
21	755.5	56.4	430	21	B58180	Lung cancer associ
22	741.5	55.3	242	21	Y67946	Human type IV coll
23	741.5	55.3	258	20	Y31892	Type IV collagen N
24	741.5	55.3	258	21	Y97554	Human alpha2(IV)NC
25	727	54.3	260	20	Y31896	Type IV collagen N
26	727	54.3	260	21	Y97558	Human alpha6(IV)NC
27	680.5	51.5	260	20	Y31894	Type IV collagen N
28	690.5	51.5	260	21	Y97556	Human alpha4(IV)NC
29	339.5	25.3	1694	19	W40109	Human alpha-6(IV)
30	334	24.9	68	21	Y95920	Human Goodpasture
31	334	24.9	72	21	Y95919	Human Goodpasture
32	334	24.9	72	21	Y95821	Human Goodpasture
33	116	8.7	21	21	Y95812	Human Goodpasture
34	113	8.4	21	21	Y95913	Human Goodpasture
35	89.5	6.7	539	11	R06647	Murine interleukin
36	85.5	6.4	409	20	Y35136	Chlamydia pneumoni
37	85.5	6.4	539	11	R07507	Murine IL-2R beta
38	85	6.3	940	20	Y34516	Porphyromonas ging
39	85	6.3	962	20	Y34589	Porphyromonas ging
40	84.5	6.3	924	19	W79192	Rat Hrs-2 polypept
41	83.5	6.2	404	15	R53767	Yeast Mel 1 protei
42	80	6.0	775	19	W79193	Human Hrs-2 partic
43	79.5	5.9	1247	16	R77548	MEK5 protein. Mu
44	79.5	5.9	1247	19	M56161	A mitogen-activate
45	79.5	5.9	1247	21	B01220	Murine MEK5. Mus

ALIGNMENTS

RESULT	1
Y67942	Y67942 standard; Protein; 245 AA.
XX	XX
AC	Y67942:
DT	03-APR-2000 (first entry)
XX	XX
DE	Human type IV collagen alpha 3 chain protein sequence SEQ ID NO:10.
XX	XX
KW	Human; type IV collagen; anti-angiogenic; angiogenesis; cancer;
KW	benign tumour; Rheumatoid arthritis; diabetic retinopathy; psoriasis;
KW	ocular angiogenesis disease; Osler-Weber Syndrome; telangiectasia;
KW	myocardial angiogenesis; plaque neovascularisation; angiodysplasia;
KW	atherosclerosis; scleroderma; hypertrophic scar; cat scratch disease;
KW	contraception; obesity.
XX	XX
OS	Homo sapiens.
XX	XX
PN	W0965940-A1.
XX	XX
PD	23-DEC-1999.
XX	XX
PF	17-JUN-1999; 99WO-US13737.
XX	XX
PR	17-JUN-1998; 98US-0089689.
PR	25-MAR-1999; 99US-0126175.
XX	XX
PA	(BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX	XX
PI	Kalluri R;
XX	XX
DR	WPI: 2000-097708/08.
DR	N-PSDB; 257158.

XX Anti-angiogenic proteins comprising the NC1 domain of the alpha 1, 2 or
 PT 3 chain of type IV collagen used in, e.g. treatment of benign tumors
 PT and rheumatoid arthritis -
 XX
 PS Claim 32; Fig 16B; 117pp; English.
 CC
 CC The present sequence represents the human type IV collagen alpha 3 chain.
 CC The present invention describes an isolated protein chosen from the NC1
 CC domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or
 CC a fragment, analogue, derivative or mutant, which has anti-angiogenic
 CC properties. The anti-angiogenic proteins, multimers and chimeras are
 CC useful for inhibiting angiogenic activity in mammalian tissue.
 CC especially for treating diseases chosen from angiogenesis-dependent
 CC cancers, benign tumours, rheumatoid arthritis, diabetic retinopathy,
 CC psoriasis, ocular angiogenesis diseases, Osler-Weber Syndrome,
 CC myocardial angiogenesis, plaque neovascularisation, telangiectasia,
 CC haemophilic joints, angiofibroma, wound granulation, intestinal
 CC adhesions, atherosclerosis, scleroderma, hypertrophic scars, cat scratch
 CC disease, Helicobacter pylori ulcers, dialysis graft vascular access
 CC stenosis, contraception and obesity. The compositions can be used to
 CC inhibit a disease characterised by angiogenic activity, in conjunction
 CC with radiation therapy, chemotherapy or immunotherapy.
 CC
 XX Sequence 245 AA:
 SQ
 Query Match 100.0%; Score 1340; DB 21; Length 245;
 Best Local Similarity 100.0%; Pred. No. 8,8e-134;
 Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 GLGKRGDSGSPATWTRGFVETRHSSOTTAIPSCPECTVPLYSGFSLFYQGNORAHGOD 60
 DB 2 GLYKRGDSGSPATWTRGFVETRHSSOTTAIPSCPECTVPLYSGFSLFYQGNORAHGOD 61
 OY 61 LGTIGSLQRTTTPLEFCNVNDVCFASRNDYSYWLSTPALMNNAPITGRALEPYIS 120
 DB 62 LGTIGSLQRTTTPLEFCNVNDVCFASRNDYSYWLSTPALMNNAPITGRALEPYIS 121
 OY 121 RCNVCEGPATAIIVHSGTPIPCPHGMISLWKGFSFIMTSGSEGTGALASPGSCLE 180
 DB 122 RCNVCEGPATAIIVHSGTPIPCPHGMISLWKGFSFIMTSGSEGTGALASPGSCLE 181
 OY 181 EFRAAPLECHGRGCTNYNSYSFWLASINPERMRKRPISPTKAGELKIIISRCOCVM 240
 DB 182 EFRAAPLECHGRGCTNYNSYSFWLASINPERMRKRPISPTKAGELKIIISRCOCVM 241
 OY 241 KKRH 244
 DB 242 KKRH 245
 DE
 XX Type IV collagen NC1 domain alpha-3 monomer.
 XX
 XX Type IV collagen; NC1 domain; non-collagenous domain; human;
 XX angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;
 XX rheumatoid arthritis; retinal neovascularization;
 XX choroidal neovascularization; macular degeneration;
 XX corneal neovascularization; retinopathy of prematurity;
 XX corneal graft rejection; neovascular glaucoma;
 XX retrolental fibroplasia; epidemic Keratoconjunctivitis;
 XX vitamin A deficiency; contact lens overwear; atopic keratitis;
 XX superior limbic keratitis; pterygium keratitis sicca; sogrens;
 XX acne rosacea; phlyctenulosis; syphilis; Mycobacteria infection;
 XX lipid degeneration; chemical burn; ulcer; herpes simplex infection;

KW Herpes zoster infection; protozoan infection; Kaposi's sarcoma;
 KW Mooren ulcer; Terrien's marginal degeneration;
 KW marginal keratolysis; trauma; systemic lupus; polyarthritis;
 KW Wegener's sarcoidosis; scleritis; Steven's Johnson disease;
 KW radial keratotomy; sickle cell anaemia; sarcoïd;
 KW pseudoxanthoma elasticum; Paget's disease; vein occlusion;
 KW artery occlusion; carotid obstructive disease; chronic uveitis;
 KW chronic vitritis; Lyme's disease; Eales disease; Bechets disease;
 KW myopia; optic pit; Stargard's disease; pars planitis;
 KW chronic retinal detachment; hyperviscosity syndrome; toxoplasmosis;
 KW post-laser complication; fibrovascular tissue proliferation;
 KW haemangioma; Osler-Weber-Rendu; AIDS; ocular neovascular disease;
 KW osteoarthritis; chronic inflammation; Crohn's disease;
 KW ulcerative colitis; psoriasis; atherosclerosis; pemphigoid.
 XX
 XX Homo sapiens.
 OS
 OS Synthetic.
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..17
 FT /note="BMA0 signal peptide"
 FT Protein 18..268
 FT /note="mature protein"
 FT Peptide 18..25
 FT /note="affinity tag"
 FT Protein 26..268
 FT /note="NC1 alpha-3 monomer"
 XX
 XX MO9949885-r2.
 XX
 XX 07-OCT-1999.
 XX
 XX 26-MAR-1999; 99WC-US06445.
 XX
 XX 27-MAR-1998; 98US-0079783.
 XX 29-OCT-1998; 98US-0106170.
 XX
 XX (UNIV) UNIV KANSAS MEDICAL CENT.
 XX
 XX Hudson BC, Sarraz MP;
 DR WPI; 1999-601297/51.
 DR N-PSDB; Z20091.
 XX
 XX Inhibition of angiogenesis with non-collagenous alpha chain monomer
 PT useful for treating e.g. tumor growth or metastasis,
 PT neovascularisation, etc
 XX
 XX Disclosure; Fig 17c; 56pp; English.
 PS
 XX This sequence represents a recombinant type IV collagen
 CC non-collagenous (NC1) domain alpha-3 polypeptide composed of a
 CC BMA0 signal sequence (which is cleaved from the mature protein) to
 CC facilitate protein secretion, and a mature protein comprising an
 CC affinity tag (facilitates purification and identification of the
 CC material) and the alpha-1 chain monomer. The invention provides
 CC methods and kits for inhibiting angiogenesis, tumour growth and
 CC metastasis, and endothelial cell interaction with the extracellular
 CC matrix, each method comprising contacting the tumour or animal
 CC tissue with 1 or more isolated type IV collagen NC1 alpha chain
 CC monomer(s) selected from the group consisting of alpha-1, alpha-2,
 CC alpha-3 and alpha-6 NC1 chain monomers (see Y31991-96). The
 CC monomers can be produced via recombinant protein expression. The
 CC polynucleotides and polypeptides are used to treat an angiogenesis-
 CC mediated disorder or condition, especially selected from solid and
 CC blood-borne tumours, diabetic retinopathy, rheumatoid arthritis,
 CC retinal neovascularization, choroidal neovascularization, macular
 CC degeneration, corneal neovascularization, retinopathy of prematurity,
 CC corneal graft rejection, neovascular glaucoma, retrolental
 CC fibroplasia, epidemic keratoconjunctivitis, vitamin A deficiency,
 CC contact lens overwear, atopic keratitis, superior limbic keratitis,
 CC pterygium keratitis sicca, sogrens, acne rosacea, phlyctenulosis,
 CC syphilis, mycobacteria infections, lipid degeneration, chemical

CC burns, bacterial ulcers, fungal ulcers, herpes simplex infections,
 CC herpes zoster infections, protozoan infections, Kaposi's sarcoma,
 CC Mooren ulcer, Terrien's marginal degeneration, marginal keratolysis,
 CC trauma, systemic lupus, polyarteritis, Wegener's sarcoidosis,
 CC scleritis, Steven's Johnson disease, radial keratotomy, sickle cell
 CC anaemia, sarcoid, pseudoxanthoma elasticum, Paget's disease, vein
 CC occlusion, artery occlusion, carotid obstructive disease, chronic
 CC uveitis, chronic uveitis, Lyme's disease, Eales disease, Behcet's
 CC disease, myopia, optic pits, stargardt disease, pars planitis,
 CC chronic retinal detachment, hyperviscosity syndromes, toxoplasmosis,
 CC post-laser complications, abnormal proliferation of fibrovascular
 CC tissue, haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular
 CC disease, osteoarthritis, chronic inflammation, Crohn's disease,
 CC ulcerative colitis, psoriasis, atherosclerosis, and pemphigoid (all
 CC claimed).

XX Sequence 268 AA;

Query Match 98.1%; Score 1314; DB 20; Length 268;
 Local Similarity 99.6%; Pred. No. 5.6e-131;
 Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 KRDSGSPATWTTGFEVTRHSOTTAIPSCPEGTVPYSGFSFLVQGNORAGODLGLTL 64
 Db 29 krdsagspatwttrgtfvttrhsqtaipscpegtvpylsgfsflvqgnqrahgqdlgtl 88
 QY 65 GSCLOQRTTMPFLFCNVNDVCFNFSRNDYSYWLSTPALMPNMAPITGRALEPYISRCTV 124
 Db 89 gscloqrtttmpflfcnvndvcfnfsrndysywlstpalmpnmmapitgralepyisrctv 148
 QY 125 CEGPAIAIAVHSQTTDIPPCPHGWSILMKGRFSIMFTSAGSEGTGOLASGSCLEEFRA 184
 Db 149 cegpaiaiahsqtdlppcphgwslmkgrfsimftsagsegtgolaspscleefra 208
 QY 185 SPFLCHGRGTCNYSNSYSFMLSINPERMRKPIPTVAGLEKIIISRCQVCMKKRH 244
 Db 209 spflechgrtcnysnsysfmlsinpermrkpiptvavaglekliisrcqvcmmkrrh 268

RESULT 3
 Y97555 ID Y97555 standard; Protein; 268 AA.
 XX

AC Y97555;
 XX
 DT 12-FEB-2001 (first entry)

Human alpha3(IV)N1 protein sequence.

KW Type IV collagen alpha chain monomer; human; inhibitor; angiogenesis;
 KW tumour growth; Integrin receptor; carcinoma; sarcoma; rhabdomyosarcoma;
 KW retinoblastoma; Ewing sarcoma; neuroblastoma; osteosarcoma; leukaemia;
 KW diabetic retinopathy; rheumatoid arthritis; neovascularisation;
 KW muscular degeneration; corneal graft rejection; vitamin A deficiency;
 KW atopic keratitis; Mycobacteria infection; chemical burn; sarcoid;
 KW Kaposi's sarcoma; sickle cell anaemia; carotid obstructive disease;
 KW chronic inflammation; psoriasis; therapy; alpha3(IV)N1.

XX Homo sapiens.
 XX
 PN MO200059532-A1.
 XX

PD 12-OCT-2000.

PF 31-MAR-2000; 2000MO-US08678.

PR 01-APR-1999; 99US-0127391.

XX (BIOS-) BIOSTRATUM INC.

XX Brooks P, Hudson B;

XX

DR WPI: 2000-664962/64.
 DR N-PSDB: A90993.

PT Use of antagonists of specific integrin receptors for inhibiting
 PT angiogenesis, tumour growth or metastases, or endothelial cell
 PT interactions with the extracellular matrix

PS Disclosure; Fig 17c; 78pp; English.

XX This sequence is a human type IV collagen alpha chain monomer,
 CC designated alpha3(IV)N1. The invention relates to a method for
 CC inhibiting angiogenesis, tumour growth or metastases, or endothelial cell
 CC interactions with the extracellular matrix, comprising contacting the
 CC cells or tissue with a polypeptide composition containing antagonists of
 CC specific integrin receptors. The methods and the antagonists are useful
 CC for inhibiting angiogenesis, tumour growth or metastases, or endothelial
 CC cell interaction with the extracellular matrix. The antagonists are also
 CC useful for treating diseases and conditions with accompanying undesired
 CC angiogenesis, e.g. solid and blood-borne tumours (e.g. melanomas,
 CC carcinomas, sarcomas, rhabdomyosarcoma, retinoblastoma, Ewing sarcoma,
 CC neuroblastoma, osteosarcoma or leukaemia). These are also applicable to
 CC treating non-tumourigenic diseases and conditions with accompanying
 CC undesired angiogenesis, e.g. diabetic retinopathy, rheumatoid arthritis,
 CC retinal neovascularisation, choroidal neovascularisation, muscular
 CC degeneration, corneal graft rejection, vitamin A deficiency, atopic
 CC keratitis, Mycobacteria infections, chemical burns, Kaposi's sarcoma,
 CC sickle cell anaemia, sarcoid, carotid obstructive disease, post-laser
 CC complications, chronic inflammation or psoriasis.

XX Sequence 268 AA;

Query Match 98.1%; Score 1314; DB 21; Length 268;
 Best Local Similarity 99.6%; Pred. No. 5.6e-131;
 Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 KRDSGSPATWTTGFEVTRHSOTTAIPSCPEGTVPYSGFSFLVQGNORAGODLGLTL 64
 Db 29 krdsagspatwttrgtfvttrhsqtaipscpegtvpylsgfsflvqgnqrahgqdlgtl 88
 QY 65 GSCLOQRTTMPFLFCNVNDVCFNFSRNDYSYWLSTPALMPNMAPITGRALEPYISRCTV 124
 Db 89 gscloqrtttmpflfcnvndvcfnfsrndysywlstpalmpnmmapitgralepyisrctv 148
 QY 125 CEGPAIAIAVHSQTTDIPPCPHGWSILMKGRFSIMFTSAGSEGTGOLASGSCLEEFRA 184
 Db 149 cegpaiaiahsqtdlppcphgwslmkgrfsimftsagsegtgolaspscleefra 208
 QY 185 SPFLCHGRGTCNYSNSYSFMLSINPERMRKPIPTVAGLEKIIISRCQVCMKKRH 244
 Db 209 spflechgrtcnysnsysfmlsinpermrkpiptvavaglekliisrcqvcmmkrrh 268

RESULT 4
 Y44171 ID Y44171 standard; Protein; 471 AA.
 XX

AC Y44171;

XX 01-FEB-2000 (first entry)

XX Bovine type IV collagen alpha3 chain protein.

KW Recombinant; bovine; alpha3 chain; type IV collagen; detection;

KW Goodpasture syndrome; antibody; blood; tissue; human; nephrolithiasis.

OS Bos taurus.

PN US5973120-A.

XX 26-OCT-1999.

PF 07-MAR-1995; 95US-0399889.

Query Match	90.3%;	Score 1210.5;	DB 20;	Length 471;
Best Local Similarity	90.6%;	Pred. No. 1.1e-119;		
Matches 221;	Conservative 10;	Mismatches 12;	Indels 1;	Gaps 1

RESULT	5
Y56783	
ID	Y56783 standard; Protein; 471 AA.

DT	27-MAR-2000	(first entry)
XX		
XX		
DE	Bovine alpha3 type IV collagen.	
XX		
XX		
KW	Goodpasture syndrome; type IV collagen; alpha3 chain; bovine	
XX		
OS		
XX	Bos sp.	
XX		
PN		
XX	US6007980-A.	
XX		
PD	28-DEC-1999.	
XX		
PF		
07-OCT-1998;	98US-0167364.	
XX		

Query Match	90.3%;	Score 1210.5;	DB 21;	Length 471;
Best Local Similarity	90.6%;	Pred. No. 1,1e-119;		
Matches 221;	Conservative 10;	Mismatches 12;	Indels 1;	Gaps 1;

RESULT	6
R79163	
ID	R79163 standard; Protein; 471 AA

DT	22-DEC-1995	(first entry)
XX		
DE	Partial sequence of bovine alpha 3 chain of type IV collagen	
XX		
KW	Type IV collagen; alpha 3 chain.	
XX		
OS	Bos taurus.	
XX		
PN	US5424408-A.	
XX		
PD	13-JUN-1995.	
XX		
PF	30-NOV-1990;	90US-0621091.
XX		
PR	30-NOV-1990;	90US-0621091.
XX		

XX WPI; 2000-096371/08.
 DR N-PSDB; 246729.
 XX
 PT Diagnosing and treating Goodpasture syndrome using a peptide derived
 PT from type IV collagen
 XX
 PS Disclosure; Columns 23-26; 26pp; English.
 XX
 CC The invention provides a method of detecting Goodpasture antibodies
 CC in the fluid of a patient by contacting it with a peptide comprising at
 CC most 218 amino acids of the human alpha 3 chain type IV collagen that
 CC contains the fragment shown in Y36785. The methods are useful for the
 CC diagnosis and treatment of Goodpasture syndrome. The present sequence
 CC represents the carboxy terminal noncollagenous domain of the human
 CC alpha3 chain of type IV collagen.
 XX
 SQ Sequence 218 AA;

Query Match 89.0%; Score 1192; DB 21; Length 218;
 Best Local Similarity 99.5%; Pred. No. 3.5e-118;
 Matches 217; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 27 OTTAIPSCPEGVPLYSGFSEFLVCGNRAHGODLGTGSCLORTTTPLEFCNVNDVCN 86
 Db 1 qttaipecegtvplysgfslfvgngnrahgqdlgtlgsclqrlttmpflfcnvndvcn 60
 QY 87 FASRNDYSYWLSTPALMPNNAPITGRALPEYISRCTVCEGPALAIIVHSQTTDIPCPH 146
 Db 61 fasnrdysywlstpalmpnnapitgralepyisrctvcgepalaiavhsqtdipcp 120
 QY 147 GMSLWKGFSFTMTSAGSEGTGALASPGSCLEFRASPLECHGRGTCNYSNSYSFW 206
 Db 121 gmslwkgfsftmtlsagsegtgalaaspyscleefrasplfchgrgtycnysnysfw 180
 QY 207 LASLNPERMFRKPIPTSTYKAGLEKTIISRCOVCMKKRH 244
 Db 181 laslnpermtkpiptstykagelkltisrcqvmckrh 218

RESULT 9
 R79164
 ID R79164 standard; Protein; 218 AA.
 XX
 AC R79164;
 XX
 DT 22-DEC-1995 (first entry)
 XX
 DB Partial sequence of human alpha 3 chain of type IV collagen.
 KW Type IV collagen; alpha 3 chain; Alport syndrome; COL4A3 gene.
 XX
 OS Homo sapiens.
 XX
 PN US5424408-A.
 XX
 PD 13-JUN-1995.
 XX
 PF 30-NOV-1990; 90US-0621091.
 XX
 PR 30-NOV-1990; 90US-0621091.
 XX
 PA (UNITV) UNIV KANSAS MEDICAL CENT.
 PA (UYTA) UNIV YALE.
 XX
 PI Hudson BG, Morrison KE, Reeders ST;
 DR WPI; 1995-262631/34.
 DR N-PSDB; 096291.
 XX
 PT cDNAs encoding human or bovine alpha-3 type 4 collagen peptide(s) -
 useful for detection and therapeutic removal of antibodies associated

PT with Goodpasture syndrome
 XX
 PS Disclosure; Columns 7-10; 33pp; English.
 XX

Using the PCR with primers derived from each end of the known 27 AA
 CC residue bovine alpha 3 (IV) collagen protein sequence, a 68 bp
 CC bovine genomic fragment was amplified. This fragment was then used
 CC to a bovine lens cDNA library and a 1.5 kb partial cDNA clone was
 CC obtd. (clone KMC15). This encodes 238 residues of the triple helical
 CC collagenous domain and all 233 residues of the C-terminal non-
 CC collagenous (NC1) domain of the alpha 3 (IV) chain. This bovine
 CC cDNA clone was used to screen a human kidney cDNA library and a
 CC 2.7 kb human cDNA clone (clone KMC27) was obtd. This clone encodes
 CC 218 residues of the NC1 domain and a portion of the 3' UTR region
 CC of the human alpha 3(IV) chain. The COL4A3 gene localises to
 CC chromosome 2 and therefore mutations in COL4A3 cannot be
 CC responsible for Alport syndrome which is X-linked. An isolated
 CC and substantially pure nt. having the sequence in 096291
 XX

Query Match 87.3%; Score 1170; DB 16; Length 218;
 Best Local Similarity 98.2%; Pred. No. 7.4e-116;
 Matches 214; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 27 OTTAIPSCPEGVPLYSGFSEFLVCGNRAHGODLGTGSCLORTTTPLEFCNVNDVCN 86
 Db 1 qttaipecegtvplysgfslfvgngnrahgqdlgtlgsclqrlttmpflfcnvndvcn 60
 QY 87 FASRNDYSYWLSTPALMPNNAPITGRALPEYISRCTVCEGPALAIIVHSQTTDIPCPH 146
 Db 61 fasnrdysywlstpalmpnnapitgralepyisrctvcgepalaiavhsqtdipcp 120
 QY 147 GMSLWKGFSFTMTSAGSEGTGALASPGSCLEFRASPLECHGRGTCNYSNSYSFW 206
 Db 121 gmslwkgfsftmtlsagsegtgalaaspyscleefrasplfchgrgtycnysnysfw 180
 QY 207 LASLNPERMFRKPIPTSTYKAGLEKTIISRCOVCMKKRH 244
 Db 181 laslnpermtkpiptstykagelkltisrcqvmckrh 218

RESULT 10
 B58169
 ID B58169 standard; Protein; 406 AA.
 XX
 AC B58169;
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Lung cancer associated polypeptide sequence SEQ ID 507.
 XX
 KW Human; Lung cancer associated protein; neuroprotective; cytostatic;
 KW cardiocactive; immunomodulatory; muscular active; vulnary;
 KW gastrointestinal; nephrotropic; antilinfecive; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease.
 XX
 OS Homo sapiens.
 XX
 PN W0200055180-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000MO-US05918.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 XX


```

RESULT 12
W09643 16-JUN-1997 (first entry)
ID W09643 standard; Protein: 772 AA.
XX W09643;
XX
XX 16-JUN-1997 (first entry)
XX Human type IV collagen alpha-5.
XX
XX Collagen alpha5(IV); basement membrane; Alport's syndrome;
XX nephritis; kidney; renal failure; antibody; diagnosis; COL4A5 gene;
XX X chromosome.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT 1..543
XX FT Domain /label= Collagenous_domain
XX FT /note= "collagenous domain contains Gly-X-Y
XX FT tripeptide repeats, interrupted at
XX FT positions 43-47, 159-160, 275-276, 334-335,
XX FT 456-459"
XX FT
XX FT Domain 544..772
XX FT /label= Non-collagenous_domain
XX FT Peptide 742..751
XX FT /label= Immunogenic_peptide
XX FT /note= "peptide used to raise diagnostic
XX FT antibodies (Claim 1)"
XX FT
XX US5593900-A.
XX
XX 14-JAN-1997.
XX
XX 07-JUL-1989; 89US-03772238.
XX
XX 20-DEC-1990; 90US-0630563.
XX 07-JUL-1989; 89US-03772238.
XX 11-OCT-1994; 94US-0321084.
XX
XX (HOST/) HOSTIKKA S L.
XX PA (HOYH/) HOYHTYA M.
XX PA (TRYG/) TRYGVASON K.
XX
XX Hostikka SL, Hoyhtya M, Trygvason K;
XX WPI: 1997-09481/09.
XX DR N-PSDB: T47812.
XX
XX New antibodies specific for human type IV collagen alpha5 chain -
XX used to detect absence of this chain in patients with renal failure
XX
XX Disclosure: Fig 2A-2B; 12pp; English.
XX
XX The amino acid sequence of a portion (W09643) of the previously
XX unknown human type IV collagen chain, alpha5(IV), was deduced from
XX cDNA clones (see also T47812) obt'd. using probes based on
XX conserved sequences of human alpha1(IV) and alpha2(IV) collagen
XX chains and of the Drosophila alpha(IV) chain. It includes a
XX complete non-collagenous domain that shows 83% identity with that
XX of alpha1(IV) and 63% with that of the alpha2(IV) chain. Mutations
XX in the alpha5(IV) gene (COL4A5) are associated with Alport's
XX syndrome. Antibodies raised against a peptide (see also W09644)
XX specific to alpha5(IV) can be used in the diagnosis of basement
XX membrane disorders such as Alport's syndrome.
XX
XX Sequence 772 AA:
XX
Query Match 70.4%; Score 943.5; DB 18; Length 772;
Best Local Similarity 67.9%; Pred. No. 3,7e-91;
Matches 165; Conservative 35; Mismatches 42; Indels 1; Gaps 1;

```

```

OY 1 CLKKRGDSGPATW-TREGVFTRHSQTATIPSCDEGTVPITSGFSFLVQGNRAHQ 59
DB 529 gpdg19pppppgtssvahnflitrlhsqtdpcpgtlyyegfsllyvgnkrahq 588
OY 60 DLGTGSCLOFRTTTPFCNVNDVCFASRNDYISWLSLSPALMPNMNAPITGRALPEYI 119
DB 589 dlgtagsclrlfstmpfcnmvncfnasrndyswlsstpepmnmqplkxgslqpti 648
OY 120 SRCVYCEPATAIAVHSOTDIPPCPHGWTSLMKGFSEFIMFTSAGSEGTQALASPGSCL 179
DB 649 srcavceapavaviahvshqlqipncpgwdslylgyfimmhtsaagegsgqalaspqsc 708
OY 180 EEPFRAPLECHGRGTQNYNSYSFWLASINPERFRKRPISVYAGELKXISQCQC 239
DB 709 eeirapliechgrgtcnyansysfwlatvdsdmsfksqselkagdlrrtrscqvc 768
OY 240 MKK 242
DB 769 mkr 771

```

```

RESULT 13
Y67943 03-APR-2000 (first entry)
ID Y67943 standard; Protein: 229 AA.
XX Y67943;
XX
XX 03-APR-2000 (first entry)
XX
XX Human type IV collagen alpha 1 chain protein sequence SEQ ID NO:2.
XX
XX Human; type IV collagen; anti-angiogenic; angiogenesis; cancer;
XX benign tumor; rheumatoid arthritis; diabetic retinopathy; psoriasis;
XX ocular angiogenesis disease; Osler-Webber Syndrome; telangiectasia;
XX myocardial angiogenesis; plaque neovascularisation; angiodiroma;
XX atherosclerosis; scleroderma; hypertrophic scar; cat scratch disease;
XX contraception; obesity.
XX
XX Homo sapiens.
XX
XX W0965940-A1.
XX
XX 23-DEC-1999.
XX
XX 17-JUN-1999; 99WO-US13737.
XX
XX 17-JUN-1998; 98US-0089689.
XX 25-MAR-1999; 99US-0126175.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
XX kalluri R;
XX
XX WPI: 2000-097708/08.
XX DR N-PSDB: 257159.
XX
XX Anti-angiogenic proteins comprising the NC1 domain of the alpha 1, 2 or
XX 3 chain of type IV collagen used in, e.g. treatment of benign tumors
XX and rheumatoid arthritis.
XX
XX Example 1; Fig 1B; 117pp; English.
XX
XX The present sequence represents the human type IV collagen alpha 1 chain.
XX The present invention describes an isolated protein chosen from the NC1
XX domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or
XX a fragment, analogue, derivative or mutant, which has anti-angiogenic
XX properties. The anti-angiogenic proteins, multimers and chimeras are
XX useful for inhibiting angiogenic activity in mammalian tissue,
XX especially for treating diseases chosen from angiogenesis-dependent
XX cancers, benign tumors, rheumatoid arthritis, diabetic retinopathy,
XX psoriasis, ocular angiogenesis diseases, Osler-Webber Syndrome,
XX myocardial angiogenesis, plaque neovascularisation, telangiectasia,
XX haemophilic joints, angiodiroma, wound granulation, intestinal

```


THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: April 13, 2001, 15:36:51 ; Search time 31.4 Seconds
(without alignments)
149.282 Million cell updates/sec

Title: US-09-543-371-10
Perfect score: 1340
Sequence: 1 GLKGRDGSPPATWTRGF.....KAGELEKTRSCQVCMKRH 244

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

1 number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents, AA: *
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep: *
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep: *
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep: *
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep: *
5: /cgn2_6/prodata/2/1aa/6CTUS_COMB.pep: *
6: /cgn2_6/prodata/2/1aa/backfills1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1210.5	90.3	471	2	US-08-399-889-24
2	1210.5	90.3	471	3	US-09-167-364-24
3	1192	89.0	218	2	US-08-399-889-25
4	1192	89.0	218	3	US-09-167-364-25
5	339.5	25.3	1694	1	US-08-494-168-2
6	83.5	6.2	404	1	US-07-602-824A-4
7	83.5	6.2	404	1	US-07-602-824A-4
8	83.5	6.2	404	1	US-07-983-451-4
9	83.5	6.2	404	1	US-08-261-578-4
10	83.5	6.2	404	1	US-08-261-577-10
11	83.5	6.2	539	6	5449755-4
12	81.5	6.1	539	6	5198359-4
13	79.5	5.9	1247	1	US-08-472-934-10
14	79.5	5.9	1247	2	US-08-323-460A-10
15	79.5	5.9	1247	2	US-08-461-146C-10
16	79.5	5.9	1247	3	US-08-461-145C-10
17	79	5.9	837	1	US-07-923-976-2
18	77.5	5.8	1248	4	US-08-882-046-6
19	76.5	5.7	783	2	US-08-545-562A-5
20	75.5	5.6	1940	2	US-08-644-271-30
21	74.5	5.6	556	3	US-08-501-572-1
22	74.5	5.6	556	3	US-09-040-444-1
23	74	5.5	154	4	US-09-193-104-12
24	72	5.4	101	1	US-07-972-387-34
25	72	5.4	101	1	US-08-431-412-34
26	72	5.4	101	1	US-08-057-971-34
27	72	5.4	101	1	US-08-057-971-81

28	71.5	5.3	616	1	US-08-385-370-2	Sequence 2, Appl
29	71.5	5.3	616	1	US-08-385-370-4	Sequence 4, Appl
30	71	5.3	254	2	US-08-475-427-1	Sequence 1, Appl
31	71	5.3	234	2	US-07-842-165-1	Sequence 1, Appl
32	71	5.3	234	4	US-08-448-398-3	Sequence 3, Appl
33	71	5.3	302	2	US-08-475-427-6	Sequence 6, Appl
34	71	5.3	302	2	US-07-842-165-6	Sequence 6, Appl
35	71	5.3	329	2	US-08-475-427-13	Sequence 13, Appl
36	71	5.3	329	2	US-07-842-165-13	Sequence 13, Appl
37	71	5.3	437	3	US-09-073-569-2	Sequence 2, Appl
38	70	5.2	101	1	US-07-972-387-26	Sequence 26, Appl
39	70	5.2	101	1	US-08-431-412-26	Sequence 26, Appl
40	70	5.2	101	1	US-08-057-971-26	Sequence 26, Appl
41	70	5.2	167	4	US-09-062-440-9	Sequence 9, Appl
42	70	5.2	574	4	US-09-062-440-2	Sequence 2, Appl
43	69.5	5.2	263	5	PCT-US94-00844-11	Sequence 11, Appl
44	69.5	5.2	364	1	US-08-318-831-6	Sequence 6, Appl
45	69.5	5.2	1333	4	US-09-356-952-2	Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-08-399-889-24
; Sequence 24, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OR INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399, 889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calif
US-08-399-889-24
```

Query Match 90.3%; Score 1210.5; DB 2; Length 471;
Best Local Similarity 90.6%; Pred. No. 1.9e-126;
Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;

```
QY 1 GLKGRDGSPPATWTR-RCFTFTRHSQTIAIPSCPECTVLYSGSFLFYQGNORAHQ 59
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 227 GLKGRDGTGPAGAVRGVFTFRHSQTIAIPSCPECTVLYSGSFLFYQGNORAHQ 286
QY 60 DLGLGSGCLQFTMPPELFECVNDVNCNFRNDYSYMLSTPALPMNNAPTTGALPEYI 119
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 287 DLGLGSGCLQFTMPPELFECVNDVNCNFRNDYSYMLSTPALPMNNAPTTGALPEYI 346
QY 120 SRCTVCEGPAIAIAVHSQTDIPPCPHGWISLWKGFSIMFTSAGSECTGQALASPGSCL 179
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 347 SRCTVCEGPAIAIAVHSQTDIPPCPHGWISLWKGFSIMFTSAGSECTGQALASPGSCL 406
QY 180 EEFASPFLEHGHTGCTNYNSYSFNLASINPRMRKPIPTVYKAGELEKTRSCQV 239
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 407 EEFASPFLEHGHTGCTNYNSYSFNLASINPRMRKPIPTVYKAGELEKTRSCQV 466
QY 240 MKKR 243
    |||
Db 467 MKKR 470
    |||
RESULT 2
US-09-167-364-24
```

```

; Sequence 24, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Calf
US-09-167-364-24

```

```

Query Match          90.3%; Score 1210.5; DB 3; Length 471;
Best Local Similarity 90.6%; Pred. No. 1.9e-126;
Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;

```

```

QY 1 GLKGRGDSSEPAWTWT-KGFVTRRSQTTATPSCPEGVPLYSGESEFLVQGNORAHQ 59
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
DB 227 GLKGRGDSSEPAWTWT-KGFVTRRSQTTATPSCPEGVPLYSGESEFLVQGNORAHQ 286
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
QY 60 DLGTLSCLORTFTMPFLFCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALPEYI 119
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
DB 287 DLGTLSCLORTFTMPFLFCNVNDVCFASRNDYSYWLSTPALPMNMAPITGRALPEYI 346
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
QY 120 SRCTVCEGPAIAIAVHSQTTDIPPCPHGWTSLMKGFSFIMFTSAGSEGTQALASPGSCL 179
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
DB 347 SRCTVCEGPAIAIAVHSQTTDIPPCPHGWTSLMKGFSFIMFTSAGSEGTQALASPGSCL 406
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
QY 180 EEFRRASPELCHRGTCNYNSYSFELASLNPBRMRKPIPTVKGELKIIISRCQVC 239
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
DB 407 EEFRRASPELCHRGTCNYNSYSFELASLNPBRMRKPIPTVKGELKIIISRCQVC 466
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
QY 240 MKRR 243
    ||| 11:1
DB 467 MKRR 470

```

```

RESULT 3
US-08-399-889-25
; Sequence 25, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399,889B
; CURRENT FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; EARLIER FILING DATE: 1990-11-30
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-08-399-889-25

```

```

Query Match          89.0%; Score 1192; DB 2; Length 218;
Best Local Similarity 99.5%; Pred. No. 7.5e-125;
Matches 217; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 27 OTTAIPSCPEGVPLYSGESEFLVQGNORAHGODLGTLSCLQRTTTPFLFCNVNDVCN 86
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
DB 1 OTTAIPSCPEGVPLYSGESEFLVQGNORAHGODLGTLSCLQRTTTPFLFCNVNDVCN 60
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
QY 87 FASRNDYSYWLSTPALPMNMAPITGRALPEYISCTVCEGPAIAIAVHSQTTDIPCPH 146
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
DB 61 FASRNDYSYWLSTPALPMNMAPITGRALPEYISCTVCEGPAIAIAVHSQTTDIPCPH 120
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
QY 147 GWISLWKGFSTIMFTSAGSEGTQALASPGSCLFEFRASPFLECHRGTCNYNSYSFWM 206
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
DB 121 GWISLWKGFSTIMFTSAGSEGTQALASPGSCLFEFRASPFLECHRGTCNYNSYSFWM 180
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
QY 207 LASLNPBRMRKPIPTVKGELKIIISRCQVCMMKRRH 244
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
DB 181 LASLNPBRMRKPIPTVKGELKIIISRCQVCMMKRRH 218
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11

```

```

RESULT 4
US-09-167-364-25
; Sequence 25, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167,364
; CURRENT FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-09-167-364-25

```

```

Query Match          89.0%; Score 1192; DB 3; Length 218;
Best Local Similarity 99.5%; Pred. No. 7.5e-125;
Matches 217; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 27 OTTAIPSCPEGVPLYSGESEFLVQGNORAHGODLGTLSCLQRTTTPFLFCNVNDVCN 86
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
DB 1 OTTAIPSCPEGVPLYSGESEFLVQGNORAHGODLGTLSCLQRTTTPFLFCNVNDVCN 60
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
QY 87 FASRNDYSYWLSTPALPMNMAPITGRALPEYISCTVCEGPAIAIAVHSQTTDIPCPH 146
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
DB 61 FASRNDYSYWLSTPALPMNMAPITGRALPEYISCTVCEGPAIAIAVHSQTTDIPCPH 120
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
QY 147 GWISLWKGFSTIMFTSAGSEGTQALASPGSCLFEFRASPFLECHRGTCNYNSYSFWM 206
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
DB 121 GWISLWKGFSTIMFTSAGSEGTQALASPGSCLFEFRASPFLECHRGTCNYNSYSFWM 180
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
QY 207 LASLNPBRMRKPIPTVKGELKIIISRCQVCMMKRRH 244
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11
DB 181 LASLNPBRMRKPIPTVKGELKIIISRCQVCMMKRRH 218
    ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11 ||||| 11:11

```

```

RESULT 5
US-08-494-168-2
; Sequence 2, Application US/08494168
; Patent No. 5731192
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T.
; APPLICANT: Zhou, Jing
; TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method
; TITLE OF INVENTION: of Detecting Collagen Deficiency
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner

```

STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/494,168
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/112,465
FILING DATE: 27-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Saxe, Bernhard D.
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 40397/104/BABR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-494-168-2

Query Match 25.3%; Score 339.5; DB 1; Length 1694;
Best Local Similarity 33.2%; Pred. No. 9,6e-29;
Matches 89; Conservative 4; Mismatches 91; Indels 47; Gaps 13;

QY 1 GLKKRGDSSP-----ATWTRGFVTRHSQTTAIP--SCPEGVPLYS 43
DB 1438 GLOGPREFEAPGQGGPSCLECLARAEMATRMSTASR-MCPRVPSASCWGKT----- 1491
QY 44 GFSELFVQGNORAHGODLTGLSCLOFRTTPEFCNVNDVCNPFASRNDYSYWLSTPALM 103
DB 1492 -AYCLMRGNTKPTTRWALLAPVCPASAPCPH-YCNINVCYHARRNDKYTWLSTPA-- 1546
QY 104 PMNAPITIGALEPIYISRCVCEGPA---IAIAVHSQTTDIPPCPHGISLWKGFSEITMF 160
DB 1547 PIPMPVPSQTOIPQYISRCVCEAPSKPFCAPQGHHPA--VPP-----GLAQPLDWLFL 1599
161 TSA---GSEGTGALASPGSCLEEFASPFLECH--GRGTCNYSNSYFPLASINPERM 215
DB 1600 PHAHCRCRCRGKROSLVSPGS---SRTGCPPLSSNAVARGTCHYANKISFWLTTVEERQ 1656
QY 216 FRK-PIPTVYAGLEKTIISRCQYCMRK 242
DB 1657 FGLPVSEITLKAGQLHTRVSRQYCMNR 1684

RESULT 6
US-07-602-824A-4
Sequence 4, Application US/07602824A
Patent No. 5356804
GENERAL INFORMATION:
APPLICANT: Desnick, Robert J.
APPLICANT: Bishop, David F.
APPLICANT: Ioannou, Yiannis A.
TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
TITLE OF INVENTION: ACTIVE alpha-GALACTOSIDASE A
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York

STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/602,824A
FILING DATE: 24-OCT-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Cortuzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 6923-005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 404 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-07-602-824A-4

Query Match 6.2%; Score 83.5; DB 1; Length 404;
Best Local Similarity 23.8%; Pred. No. 0.4;
Matches 35; Conservative 23; Mismatches 58; Indels 31; Gaps 7;

QY 1 GLKKRGDSSPATTWTRGFV--FTRHSQTTAIPSCPEGVPLYSGFSELFVQGNORAH 57
DB 181 GYV---SGIANSWRMSGDYTAFTTRPD-----SCPDG---YYAGFSIMILKKAAPM 227
QY 58 GODCLTGLSCLOFRTTPEFCNVNDVCNPFASRNDYSYWLSTPALMNMNAPITIGAL 115
DB 228 GQMGVVG-----WNDDLNLEVGGNLTDDEKAFHSMMAMVKSPLIGAVNNLKASSY 282
QY 116 EPIYISRCVCEGPAIAIAVHSQTTDIP 142
DB 283 SIY-----SQASVIAINDSNGIP 301

RESULT 7
US-07-602-608-4
Sequence 4, Application US/07602608
Patent No. 5382524
GENERAL INFORMATION:
APPLICANT: Desnick, Robert J.
APPLICANT: Bishop, David F.
APPLICANT: Ioannou, Yiannis A.
APPLICANT: Wang, Anne M.
TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
TITLE OF INVENTION: ACTIVE ALPHA-N-ACETYLALACTOSAMINIDASE
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/602,608

FILED DATE: 24-OCT-1990
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 6923-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 404 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-07-602-608-4

Query Match 6.2%; Score 83.5; DB 1; Length 404;
Best Local Similarity 23.8%; Pred. No. 0.4;
Matches 35; Conservative 23; Mismatches 58; Indels 31; Gaps 7;

QY 1 GLGKRDGSGPATMTTRGFV---FTRHSQTALPSCPEGTPLYSGFSLFVQGNORAH 57
DB 181 GLYG---SGIANSWRMSGDVTAETRPD-----SCPDG---YTAGFSIMNLKAPM 227
QY 58 GODLGLTGLSCLOFTTTPFLFCNVNDVCNPFASRNDYSYV--LSTPALPMNMAPITGRAL 115
DB 228 GQNAVGCG---WNDLDNLEVGVCNLTDEDEKAFHSMWAMVKSPLIGANVNNLKASSY 282
QY 116 EPIISRCTVCEGPALAIYVHSQTTDIP 142
DB 283 SIY-----SQASVYAIINQDSNGIP 301

RESULT 8
US-07-983-451-4
Sequence 4, Application US/07983451
Patent No. 5401650
GENERAL INFORMATION:
APPLICANT: Desnick, Robert J.
APPLICANT: Bishop, David F.
APPLICANT: Ioannou, Yiannis A.
TITLE OF INVENTION: Cloning and Expression of Biologically
TITLE OF INVENTION: Active alpha-galactosidase A
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/983,451
FILING DATE: 30-NOV-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 07/983,451
REFERENCE/DOCKET NUMBER: 6923-030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 404 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-07-983-451-4

Query Match 6.2%; Score 83.5; DB 1; Length 404;
Best Local Similarity 23.8%; Pred. No. 0.4;
Matches 35; Conservative 23; Mismatches 58; Indels 31; Gaps 7;

QY 1 GLGKRDGSGPATMTTRGFV---FTRHSQTALPSCPEGTPLYSGFSLFVQGNORAH 57
DB 181 GLYG---SGIANSWRMSGDVTAETRPD-----SCPDG---YTAGFSIMNLKAPM 227
QY 58 GODLGLTGLSCLOFTTTPFLFCNVNDVCNPFASRNDYSYV--LSTPALPMNMAPITGRAL 115
DB 228 GQNAVGCG---WNDLDNLEVGVCNLTDEDEKAFHSMWAMVKSPLIGANVNNLKASSY 282
QY 116 EPIISRCTVCEGPALAIYVHSQTTDIP 142
DB 283 SIY-----SQASVYAIINQDSNGIP 301

RESULT 9
US-08-261-578-4
Sequence 4, Application US/08261578
Patent No. 5491075
GENERAL INFORMATION:
APPLICANT: Desnick, Robert J.
APPLICANT: Bishop, David F.
APPLICANT: Ioannou, Yiannis A.
APPLICANT: Wang, Anne M.
TITLE OF INVENTION: CLONING AND EXPRESSION OF BIOLOGICALLY
TITLE OF INVENTION: ACTIVE ALPHA-N-ACETYLGLACTOSAMINIDASE
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/261,578
FILING DATE: 17-JUN-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/602,608
FILING DATE: 24-OCT-1990
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 6923-008
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 404 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-261-578-4


```

OY 48 LPVQGNQRAHGDDIGTACSLQREFTTIPPELFQVANDVCMFAR-----NYSWLSTPA 100
    ::: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 264 VYLLVNCRYLGPMLKTV-----LKCHPPSEFESQLSQHGGDLQKMWSP 310
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 102 LPMNNAPITGRLREYXSRCVCEGPAIAVAHSOTDIPCP--HGWLISW-KGF 155
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 311 -VPLSFSPSGPR--PEISPLEVLDGDSKAVQLLLQKDSAPLPSPSGHSQASCFPTNGY 367
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 156 SFIMFTSAGSEGTGQALSPGSCLEE 181
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 368 FFHLHPALALESCQYVTFYDPCVEE 393
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 13
US-08-472-

Sequence 10, Application US/08472934
Patent No. 5753446
GENERAL INFORMATION:
APPLICANT: JOHNSON, GARY L.
TITLE OF INVENTION: METHOD AND PRODUCT FOR REGULATING CELL
TITLE OF INVENTION: RESPONSIVENESS TO EXTERNAL SIGNALS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lahive and Cockfield
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,934
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/440,421
FILING DATE: 15-May-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/354,516
FILING DATE: 21-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 5,405,941
FILING DATE: 15-Apr-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/323,460
FILING DATE: 14-Oct-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/11690
FILING DATE: 14-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04178
FILING DATE: 15-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Giulio A. Decontili, Jr. Esq.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: CFI-004DVC2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1247 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-472-934-10

Query Match 5.9% Score 79.5; DB 1; Length 1247

[illegible]

RESULT 14
ME-08-323-

US-08-323-460A-10
Sequence 10; Application US/08323460A
Patent No. 5854043
GENERAL INFORMATION:
APPLICANT: JOHNSON, GARY L.
TITLE OF INVENTION: METHOD AND PRODUCT FOR REGULATING CELL
TITLE OF INVENTION: RESPONSIVENESS TO EXTERNAL SIGNALS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: SHERIDAN ROSS & MCINTOSH
STREET: 1700 LINCOLN STREET, SUITE 3500
CITY: DENVER
STATE: CO
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,460A
FILING DATE: 14-OCT-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/049,254
FILING DATE: 14-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: KOVARIK, JOSEPH E.
REGISTRATION NUMBER: 33,005
REFERENCE/DOCKET NUMBER: 2879-1-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 303/863-9700
TELEFAX: 303/863-0223
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1247 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

```

Query Match          5.9%; Score 79.5; DB 2; Length 1247;
Best Local Similarity 23.0%; Pred. No. 5.4;
Matches 51; Conservative 17; Mismatches 89; Indels 75; Gaps 9
OY      10 GSPATWTTGCFVFYTR-----HSQTTAIG----- 31
      1: 1 1 1 1 1:
Db      727 GTRPRMATAQGFDELAIEPAFISALPEDDFLSLQALMNECGHYIGKPHSYVTALHHNSP 786
      32 -----PSC-----PECTVPLVYGGSFSLPVQGNORAHAGODLGTGSLQRTFTTPPLFCN 80

```

Db 787 RPVKVPRCHSDPPNPPLIIPTEGFSRVSVDARTHCNSVAAAARAATTA----- 840
QY 81 VNDVCNFSRNDYSWLSSTPALPMNMAPITGRALEPIYSRCT--VCEGPATAIAVHSOT 138
Db 841 -----AGRPGPGGDSVPA-KPVNTAPDTRGSSVPENDRLASIAELQFSLSRHSSP 892
QY 139 T---DIPCPHGWMIS-----LWKGFSFIMFT--SAGSEGTGALASPGSCLEEFR 183
Db 893 TERDEPAPVPRSDSSGSTRSRWELRTLISQTKDSASKOGPIEAIQKSVRLFEEER 947

RESULT 15
US-08-461-146C-10
Sequence 10. Application US/08461146C
Patent No. 5981265
GENERAL INFORMATION:
APPLICANT: JOHNSON, GARY L.
TITLE OF INVENTION: METHODS FOR REGULATING MEK PROTEIN ACTIVITY
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lahive and Cockfield
STREET: 60 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: U.S.A.
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/461,146C
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/354,516
FILING DATE: 21-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/049,254
FILING DATE: 15-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/323,460
FILING DATE: 14-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/11690
FILING DATE: 14-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/04178
FILING DATE: 15-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: KARA, Catherine J.
REGISTRATION NUMBER: P41,106
REFERENCE/DOCKET NUMBER: CPI-004CN3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1247 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-461-146C-10

Query Match 5.9%; Score 79.5; DB 2; Length 1247;
Best Local Similarity 23.0%; Pred. No. 5.4;
Matches 54; Conservative 17; Mismatches 89; Indels 75; Gaps 9;

QY 10 GSPATWTTTGGFVETR-----HSQTTAI----- 31
| : - | : | | | : |
| | | | |

Db 727 GTRPRMATOGFDFLOAIEPAFISALPEDDFLSLOALMNECIGHVIGKPHSPVTAIHRNSP 786
QY 32 -----PSC-----PEGIVPLISGFSFLFVQGNQRAHGODLITGSCLOFTTMTPLPCN 80
Db 787 RPVKVPRCHSDPPNPPLIIPTEGFSRVSVDARTHCNSVAAAARAATTA----- 840
QY 81 VNDVCNFSRNDYSWLSSTPALPMNMAPITGRALEPIYSRCT--VCEGPATAIAVHSOT 138
Db 841 -----AGRPGPGGDSVPA-KPVNTAPDTRGSSVPENDRLASIAELQFSLSRHSSP 892
QY 139 T---DIPCPHGWMIS-----LWKGFSFIMFT--SAGSEGTGALASPGSCLEEFR 183
Db 893 TERDEPAPVPRSDSSGSTRSRWELRTLISQTKDSASKOGPIEAIQKSVRLFEEER 947

Search completed: April 13, 2001, 15:41:57
Job time: 306 sec

THIS PAGE BLANK (USPTO)

DB 182 EFRASPFLCHGRCNTNYSNSYSFWLASLNPFRMRKPIPTVKGAGLEKTIISRCQVCM 241
 OY 241 KKRH 244
 DB 242 KKRH 245

RESULT 2
 ID 061435 PRELIMINARY; PRT; 246 AA.

AC 061435;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)
 DE PROCOLLAGEN TYPE IV, ALPHA 3 (COLLAGEN IV ALPHA 3 CHAIN) (FRAGMENT).
 GN COL4A3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;

SEQUENCE FROM N.A.

RC STRAIN=BAIB/C;
 MEDLINE=95050957; PubMed=7962065;
 RA Miner J.H., Sanes J.R.;
 RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
 laminae: sequence, distribution, association with laminins, and
 developmental switches.";
 RL J. Cell Biol. 127:879-891(1994).
 RN (12)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BAIB/C;
 RA Miner J.H.;
 RL Submitted (JUL-1994) to the EMBL/Genbank/DBJ databases.

SEQUENCE FROM N.A.

RC STRAIN=BAIB/C;
 RA Miner J.H.;
 RL Submitted (JUL-1994) to the EMBL/Genbank/DBJ databases.

SEQUENCE FROM N.A.

RC STRAIN=BAIB/C;
 RA Miner J.H.;
 RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
 DR EMBL: 235166; CA84529.1;
 DR MGD: MGI:104688; Col4a3.
 DR INTERPRO: IPR000504; -;
 DR INTERPRO: IPR001442; -;
 DR PFIAM: PFI01413; C4; 2.
 DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
 FT NON TER
 SO SEQUENCE 246 AA; 26993 MW; A9B5434F5836F324 CRC64;

Query Match 91.2%; Score 1221.5; DB 11; Length 246;
 Best Local Similarity 90.6%; Pred. No. 6, 9e-115;
 Matches 222; Conservative 11; Mismatches 11; Indels 1; Gaps 1;

OY 1 GUKGRGDSGSPATWT-TGCEVFTRHSTQTTAIPSCCEGTVPVLYSGSFLEVOGNRAHQ 59

DB 2 GUKGNPDGDTGTPATGTRMGFTFTRHSTQTTAIPSCCEGTQPIYSGSFLFVGNKRAHQ 61

OY 60 DIGTLAGSCLORFETMPFLPCNVNDVCFASRNDYSYWLSTPALMPNMAPISGRALPEYI 119

DB 62 DIGTLAGSCLORFETMPFLPCNVNDVCFASRNDYSYWLSTPALMPNMAPISGRALPEYI 121

OY 120 SRCTVCEGPALIAVHSQTTDIPPCPHGWIISLKGFSFIIMFTSAGSEGGQALASPGSCL 179

DB 122 SRCTVCEGPALIAVHSQTTAIPPCQDWSLKGFSFIIMFTSAGSEGGQALASPGSCL 181

OY 180 EEFRAAPFLCHGRCNTNYSNSYSFWLASLNPFRMRKPIPTVKGAGLEKTIISRCQVC 239

DB 182 EEFRAAPFLCHGRCNTNYSNSYSFWLASLNPFRMRKPIPTVKGAGLEKTIISRCQVC 241

OY 240 MKRH 244

DB 242 MKRH 246

RESULT 3
 ID 090ZS0 PRELIMINARY; PRT; 1669 AA.

AC 090ZS0;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, last annotation update)
 DE ALPHA 3 COLLAGEN IV.
 GN COL4A3.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;

SEQUENCE FROM N.A.

RC TISSUE=KIDNEY;
 RX MEDLINE=20005934; PubMed=10534397;
 RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,
 RA Elder F.F.B., Miner J.H., Overbeek P.A., Weisler M.H.;
 RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a
 mouse model of alport syndrome.";
 RL Genomics 61:113-124(1999).
 DR EMBL: AF169387; AAD50449.1;
 DR MGD: MGI:104688; Col4a3.
 DR INTERPRO: IPR000087; -;
 DR INTERPRO: IPR000504; -;
 DR INTERPRO: IPR001442; -;
 DR PFIAM: PFI01413; C4; 2.
 DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
 KM Collagen.
 SO SEQUENCE 1669 AA; 161769 MW; 30976E59739A47B2 CRC64;

Query Match 91.2%; Score 1221.5; DB 11; Length 1669;
 Best Local Similarity 90.6%; Pred. No. 5, 8e-114;
 Matches 222; Conservative 11; Mismatches 11; Indels 1; Gaps 1;

OY 1 GUKGRGDSGSPATWT-TGCEVFTRHSTQTTAIPSCCEGTVPVLYSGSFLEVOGNRAHQ 59

DB 1425 GUKGNPDGDTGTPATGTRMGFTFTRHSTQTTAIPSCCEGTQPIYSGSFLFVGNKRAHQ 1484

OY 60 DIGTLAGSCLORFETMPFLPCNVNDVCFASRNDYSYWLSTPALMPNMAPISGRALPEYI 119

DB 1485 DIGTLAGSCLORFETMPFLPCNVNDVCFASRNDYSYWLSTPALMPNMAPISGRALPEYI 1544

OY 120 SRCTVCEGPALIAVHSQTTDIPPCPHGWIISLKGFSFIIMFTSAGSEGGQALASPGSCL 179

DB 1545 SRCTVCEGPALIAVHSQTTAIPPCQDWSLKGFSFIIMFTSAGSEGGQALASPGSCL 1604

OY 180 EEFRAAPFLCHGRCNTNYSNSYSFWLASLNPFRMRKPIPTVKGAGLEKTIISRCQVC 239

DB 1605 EEFRAAPFLCHGRCNTNYSNSYSFWLASLNPFRMRKPIPTVKGAGLEKTIISRCQVC 1664

OY 240 MKRH 244

DB 1665 MKRH 1669

RESULT 4

ID 063122 PRELIMINARY; PRT; 230 AA.

AC 063122;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, last annotation update)
 DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).
 GN COL4A3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;

SEQUENCE FROM N.A.

RC STRAIN=BAIB/C;
 RA Miner J.H., Sanes J.R.;
 RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
 laminae: sequence, distribution, association with laminins, and
 developmental switches.";
 RL J. Cell Biol. 127:879-891(1994).
 RN (12)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BAIB/C;
 RA Miner J.H.;
 RL Submitted (JUL-1994) to the EMBL/Genbank/DBJ databases.

SEQUENCE FROM N.A.

RC STRAIN=BAIB/C;
 RA Miner J.H.;
 RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.

RC STRAIN=BAIB/C;
 RA Miner J.H.;
 RL Submitted (JUL-1994) to the EMBL/Genbank/DBJ databases.

RP SEQUENCE FROM N.A.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE=KIDNEY CORTEX;
 RA Turner N., Ryan J.J., Derry C.J., Cashman S.J., Kathanna I.,
 RA Mason P.J., Pusey C.D.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: I47281; AAB72238.2; -
 DR INTERPRO: IPR000504; -
 DR INTERPRO: IPR001442; -
 DR PRAM: PF01413; C4; 2;
 DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 230
 SO SEQUENCE 230 AA; 25398 MW; 29549E25314CC056 CRC64;

Query Match 86.9%; Score 1164; DB 11; Length 230;
 Best Local Similarity 91.6%; Pred. No. 3.8e-109;
 Matches 208; Conservative 11; Mismatches 8; Indels 0; Gaps 0;

Q 18 RGFFTRHSOTTAIPSCPGTVPVLYSGFSLFVQGNQRAHGGDGLTGLSCLQRTTMTPEL 77
 4 RGFFTRHSOTTAIPSCPGTVPVLYSGFSLFVQGNQRAHGGDGLTGLSCLQRTTMTPEL 63
 QY 78 FCNVNDVCFNARSNDYSYMLSTPALMPNMNAPITGRALPEYISRCTVCEGPAIAVHSQ 137
 DB 64 FCNVNDVCFNARSNDYSYMLSTPALMPNMNAPITGRALPEYISRCTVCEGPAIAVHSQ 123
 QY 138 TTPDIPCPHGMISLWKGFSTMTSAGSEGTGQALASPGSCLEEFRAAPFLCHGRTCN 197
 DB 124 TTAIPPCDQGVSLMKGFSTMTSAGSEGTGQALASPGSCLEEFRAAPFLCHGRTCN 183
 QY 198 YSNSYSFWMLASLNPBRFRKPIPTVKAGLEKIISRCOVCMKKR 244
 DB 184 YSNSYSFWMLASLNPBRFRKPIPTVKAGLEKIISRCOVCMKKR 230

RESULT 5
 ID Q28512 PRELIMINARY; PRT; 212 AA.
 AC Q28512;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DE 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).
 GN COL4A3.
 OS Macaca mulatta (Rhesus macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecoidea; Macaca.
 NCBI_Taxid=9544;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=RENAL CORTEX;
 RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Kathanna I.,
 RA Mason P.J., Pusey C.D.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: I47280; AAA91861.1; -
 DR INTERPRO: IPR000504; -
 DR INTERPRO: IPR001442; -
 DR PRAM: PF01413; C4; 2;
 DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 212
 SO SEQUENCE 212 AA; 23469 MW; 4BC574A64E357E64 CRC64;

Query Match 86.4%; Score 1158; DB 6; Length 212;
 Best Local Similarity 99.1%; Pred. No. 1.4e-108;
 Matches 210; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 33 SCPEGTVPLVSGFSLFVQGNQRAHGGDGLTGLSCLQRTTMTPELFCNVNDVCFNARSND 92
 DB 1 SCPEGTVPLVSGFSLFVQGNQRAHGGDGLTGLSCLQRTTMTPELFCNVNDVCFNARSND 60

QY 93 YSYWLSTPALMPNMNAPITGRALPEYISRCTVCEGPAIAVHSQTDIPCPHGMISLW 152
 DB 61 YSYWLSTPALMPNMNAPITGRALPEYISRCTVCEGPAIAVHSQTDIPCPHGMISLW 120
 QY 153 KGFSFIMFTSAGSEGTGQALASPGSCLEEFRAAPFLCHGRTCNYSNSYSFWMASLNP 212
 DB 121 KGFSFIMFTSAGSEGTGQALASPGSCLEEFRAAPFLCHGRTCNYSNSYSFWMASLNP 180
 QY 213 ERMRKPIPTVKAGLEKIISRCOVCMKKR 244
 DB 181 ERMRKPIPTVKAGLEKIISRCOVCMKKR 212

RESULT 6
 ID Q28567 PRELIMINARY; PRT; 212 AA.
 AC Q28567;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).
 GN COL4A3.
 OS Ovis aries (Sheep).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Ovis.
 NCBI_Taxid=9940;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=RENAL CORTEX;
 RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Kathanna I.,
 RA Mason P.J., Pusey C.D.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: I47282; AAA91904.1; -
 DR INTERPRO: IPR000504; -
 DR INTERPRO: IPR001442; -
 DR PRAM: PF01413; C4; 2;
 DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 212
 SO SEQUENCE 212 AA; 23417 MW; 0F5839FCH81BDD8C CRC64;

Query Match 82.3%; Score 1103; DB 6; Length 212;
 Best Local Similarity 92.4%; Pred. No. 4.7e-103;
 Matches 195; Conservative 13; Mismatches 3; Indels 0; Gaps 0;

QY 33 SCPEGTVPLVSGFSLFVQGNQRAHGGDGLTGLSCLQRTTMTPELFCNVNDVCFNARSND 92
 DB 1 SCPEGTVPLVSGFSLFVQGNQRAHGGDGLTGLSCLQRTTMTPELFCNVNDVCFNARSND 60
 QY 93 YSYWLSTPALMPNMNAPITGRALPEYISRCTVCEGPAIAVHSQTDIPCPHGMISLW 152
 DB 61 YSYWLSTPALMPNMNAPITGRALPEYISRCTVCEGPAIAVHSQTDIPCPHGMISLW 120
 QY 153 KGFSFIMFTSAGSEGTGQALASPGSCLEEFRAAPFLCHGRTCNYSNSYSFWMASLNP 212
 DB 121 KGFSFIMFTSAGSEGTGQALASPGSCLEEFRAAPFLCHGRTCNYSNSYSFWMASLNP 180
 QY 213 ERMRKPIPTVKAGLEKIISRCOVCMKKR 243
 DB 181 ERMRKPIPTVKAGLEKIISRCOVCMKKR 211

RESULT 7
 ID Q28273 PRELIMINARY; PRT; 210 AA.
 AC Q28273;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE COLLAGEN TYPE IV ALPHA 3 CHAIN (FRAGMENT).

GN COL4A3.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 NCBI_TaxID=9615;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=SAWOED;
 RX MEDLINE=96278820; PubMed=8662866;
 RA Thorner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
 RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
 of collagen type IV. Evidence from a canine model of X-linked
 nephritis with a COL4A5 gene mutation."
 RL J. Biol. Chem. 271:13821-13828(1996).
 DR EMBL; U50935; AAC48585.1; -
 DR INTERPRO; IPR000504; -
 DR INTERPRO; IPR001442; -
 DR PFAM; PF01413; C4; 2.
 DR PROSITE; PS00030; RNP_1; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 210
 SEQUENCE 210 AA; 23025 MW; 31119EACAB23633D CRC64;

Query Match 80.9%; Score 1084; DB 6; Length 210;
 Best Local Similarity 91.9%; Pred. No. 3.7e-101;
 Matches 193; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

QY 23 TRISQTTAIPSCGEGTVPVLYSGSFLEVOGNORAHGODLGTGSCLORTTMTPELCNNV 82
 DB 1 TRISQTTAIPSCGEGTVPVLYSGSFLEVOGNORAHGODLGTGSCLORTTMTPELCNNV 60
 QY 83 DVCCNFASRNDYSYWLSTPALMPNNMAPITGRALPEYISRCTVCEGPAIAVHSQTTDIP 142
 DB 61 NVCCNFASRNDYSYWLSTPALMPNNMAPITGRALPEYISRCTVCEGPAIAVHSQTTDIP 120
 QY 143 PCPHGISTLMKGSFTMTFSAGSEGTGQALASGSCLEFRASPFTECHGRTGTCNYSNS 202
 DB 121 SCPNGISITLMKGSFTMTFSAGSEGTGQALASGSCLEFRASPFTECHGRTGTCNYSNS 180
 QY 203 YSFWLASLNPDRFRKPISTYKAGLEKTI 232
 DB 181 YSFWLASLNPDRFRKPISTYKAGLEKTI 210

RESULT 8
 Q28682 PRELIMINARY: PRT; 203 AA.
 AC Q28682;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).
 GN COL4A3
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 NCBI_TaxID=9986;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP TISSUE=RENAL CORTEX;
 RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbama I.,
 RA Mason P.J., Pusey C.D.;
 RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
 DR EMBL; I47283; AAA91893.1; -
 DR INTERPRO; IPR000504; -
 DR INTERPRO; IPR001442; -
 DR PFAM; PF01413; C4; 2.
 DR PROSITE; PS00030; RNP_1; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 203
 SEQUENCE 203 AA; 22213 MW; E14173816E4D9E30 CRC64;

Query Match 80.0%; Score 1072; DB 6; Length 203;
 Best Local Similarity 95.1%; Pred. No. 5.8e-100;
 Matches 193; Conservative 7; Mismatches 3; Indels 0; Gaps 0;
 QY 33 SCPEGTVPLYSGFSFLFVOGNORAHGODLGTGSCLORTTMTPELCNNVNDVCFASRND 92
 DB 1 SCPEGTVPLYSGFSFLFVOGNORAHGODLGTGSCLORTTMTPELCNNVNDVCFASRND 60
 QY 93 YSWLSTPALMPNNMAPITGRALPEYISRCTVCEGPAIAVHSQTTDIPPCPHGISLM 152
 DB 61 YSWLSTPALMPNNMAPITGRALPEYISRCTVCEGPAIAVHSQTTDIPPCPHGISLM 120
 QY 153 KGSFTMTFSAGSEGTGQALASGSCLEFRASPFTECHGRTGTCNYSNSYSFWLASLNP 212
 DB 121 KGSFTMTFSAGSEGTGQALASGSCLEFRASPFTECHGRTGTCNYSNSYSFWLASLNP 180
 QY 213 ERMFRKPISTYKAGLEKTIIS 235
 DB 181 ERMFRKPISTYKAGLEKTIIS 203

RESULT 9
 Q29032 PRELIMINARY: PRT; 203 AA.
 AC Q29032;
 DT 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).
 GN COL4A3.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 NCBI_TaxID=9823;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP TISSUE=RENAL CORTEX;
 RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbama I.,
 RA Mason P.J., Pusey C.D.;
 RL Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.
 DR EMBL; I47284; AAA91882.1; -
 DR INTERPRO; IPR000504; -
 DR INTERPRO; IPR001442; -
 DR PFAM; PF01413; C4; 2.
 DR PROSITE; PS00030; RNP_1; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 203
 SEQUENCE 203 AA; 22326 MW; E3B20E32D9A245AA CRC64;

Query Match 78.7%; Score 1055; DB 6; Length 203;
 Best Local Similarity 93.1%; Pred. No. 2.9e-98;
 Matches 189; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

QY 33 SCPEGTVPLYSGFSFLFVOGNORAHGODLGTGSCLORTTMTPELCNNVNDVCFASRND 92
 DB 1 SCPEGTVPLYSGFSFLFVOGNORAHGODLGTGSCLORTTMTPELCNNVNDVCFASRND 60
 QY 93 YSWLSTPALMPNNMAPITGRALPEYISRCTVCEGPAIAVHSQTTDIPPCPHGISLM 152
 DB 61 YSWLSTPALMPNNMAPITGRALPEYISRCTVCEGPAIAVHSQTTDIPPCPHGISLM 120
 QY 153 KGSFTMTFSAGSEGTGQALASGSCLEFRASPFTECHGRTGTCNYSNSYSFWLASLNP 212
 DB 121 KGSFTMTFSAGSEGTGQALASGSCLEFRASPFTECHGRTGTCNYSNSYSFWLASLNP 180
 QY 213 ERMFRKPISTYKAGLEKTIIS 235
 DB 181 ERMFRKPISTYKAGLEKTIIS 203

RESULT 10

0919K3 PRELIMINARY; PRT; 979 AA.
 ID 0919K3
 AC 0919K3
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE COLLAGEN IV A1 CHAIN (FRAGMENT).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 NCBI_TaxID=9031;
 RN
 RP SEQUENCE FROM N.A.
 RA Halfter W.M., Dong S.:
 RT "Composition, synthesis and assembly of the embryonic chick retinal
 basal lamina".
 RL Dev. Biol. 0:0-0(2000).
 DR EMBL; AF239838; AAF44681.1;
 FT NON_TER 1
 S SEQUENCE 979 AA; 95020 MW; 5B1017D911ED4299 CRC64;

Query Match 71.5%; Score 957.5; DB 13; Length 979;
 Best Local Similarity 67.3%; Pred. No. 1e-87;
 Matches 167; Conservative 33; Mismatches 41; Indels 7; Gaps 1;

QY 1 GLKGRDSDSP-----ATWTRGFVFTRHSGTTPALPSCPESTVPLYSGFSLFVQGN 53
 DB 730 GIPGAPPEGLPGAMGPPGASVAHGLVTRHSQTTEPQCPGTRILYHGYSLLYVQGN 789
 QY 54 QRAHGDGLTGLSCLQRTFTMPFLFCVNDVNCNPNASNDYSYWLSTPALMNNAPITGR 113
 DB 790 ERAHGDGLTGLSCLQRTFTMPFLFCVNDVNCNPNASNDYSYWLSTPALMNNAPITGE 849
 QY 114 ALEPISRCTVCEGPALAIIVHSOTTDIPCPHGMISLMKGFSTIMFTSAGSEGTGALA 173
 DB 850 SIRFISRCSCVCEAPAVIAVHSOTIIPCPHGMISLMKGFSTIMFTSAGSEGTGALA 909
 QY 174 SPGSCLEEFASPLECHGRTCTNYNSYSFWLASLNPENMFRRKPIPTVKAGLEKII 233
 DB 910 SPGSCLEEFASPLECHGRTCTNYNSYSFWLASLNPENMFRRKPIPTVKAGLEKII 969
 QY 234 SRCVCKM 241
 DB 970 SRCVCKM 977
 QY 11
 DB 061436 PRELIMINARY; PRT; 253 AA.
 AC 061436
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
 DE PROCOLLAGEN, TYPE IV, ALPHA 5 (COLLAGEN IV ALPHA 5 CHAIN) (FRAGMENT).
 OS COL4A5.
 OC Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 RN
 RP SEQUENCE FROM N.A.
 RA Tissue=MUSCLE;
 RX MEDLINE=95050957; PubMed=7962065;
 RA Miner J.H., Sames J.R.:
 RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
 RT laminae: sequence, distribution, association with laminins, and
 RT developmental switches".
 RL J. Cell Biol. 127:879-891(1994).
 DR EMBL; Z35168; CAA84531.1;
 DR MGD; MGI:88456; Col4a5.
 DR INTERPRO; IPR000087;

DR INTERPRO; IPR001442;
 DR PFAM; PF01413; C4; 2.
 FT NON_TER 1
 S SEQUENCE 253 AA; 27626 MW; 33DA1190CA59FA91 CRC64;

Query Match 71.1%; Score 952.5; DB 11; Length 253;
 Best Local Similarity 68.7%; Pred. No. 7.4e-88;
 Matches 167; Conservative 32; Mismatches 43; Indels 1; Gaps 1;

QY 1 GLKGRDSDSPATWT-TRGFVFTRHSGTTPALPSCPESTVPLYSGFSLFVQGNRAHQ 59
 DB 10 GPDGLOGPPGPGPTSSVAHGLFLTRHSQTTEAPQCPGCTVHIEGFSLLYVQGNRAHQ 69
 QY 60 DLGTLGSLQRTFTMPFLFCVNDVNCNPNASNDYSYWLSTPALMNNAPITGRALPYI 119
 DB 70 DLGTLGSLQRTFTMPFLFCVNDVNCNPNASNDYSYWLSTPALMNNAPITGRALPYI 129
 QY 120 SRCVCEGPALAIIVHSOTTDIPCPHGMISLMKGFSTIMFTSAGSEGTGALAPSGCL 179
 DB 130 SRCVCEGPALAIIVHSOTTDIPCPHGMISLMKGFSTIMFTSAGSEGTGALAPSGCL 189
 QY 180 EEFASPLECHGRTCTNYNSYSFWLASLNPENMFRRKPIPTVKAGLEKIIISRCVC 239
 DB 190 EEFASPLECHGRTCTNYNSYSFWLASLNPENMFRRKPIPTVKAGLEKIIISRCVC 249
 QY 240 MKK 242
 DB 250 MKR 252

RESULT 12

QY 09NDB7 PRELIMINARY; PRT; 886 AA.
 ID 09NDB7
 AC 09NDB7
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE DA24A23.1 (COLLAGEN, TYPE IV, ALPHA 5 (ALPRT SYNDROME)) (FRAGMENT).
 OS COL4A5.
 OC Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN
 RP SEQUENCE FROM N.A.
 RA Codley V.;
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL035425; CAB90289.1;
 KW Collagen.
 FT NON_TER 1
 S SEQUENCE 886 AA; 83478 MW; 8C06B9FCA9AA6569 CRC64;

Query Match 70.4%; Score 943.5; DB 4; Length 886;
 Best Local Similarity 67.9%; Pred. No. 2.4e-86;
 Matches 165; Conservative 35; Mismatches 42; Indels 1; Gaps 1;

QY 1 GLKGRDSDSPATWT-TRGFVFTRHSGTTPALPSCPESTVPLYSGFSLFVQGNRAHQ 59
 DB 643 GPDGLOGPPGPGPTSSVAHGLFLTRHSQTTEAPQCPGCTVHIEGFSLLYVQGNRAHQ 702
 QY 60 DLGTLGSLQRTFTMPFLFCVNDVNCNPNASNDYSYWLSTPALMNNAPITGRALPYI 119
 DB 703 DLGTLGSLQRTFTMPFLFCVNDVNCNPNASNDYSYWLSTPALMNNAPITGRALPYI 129
 QY 120 SRCVCEGPALAIIVHSOTTDIPCPHGMISLMKGFSTIMFTSAGSEGTGALAPSGCL 179
 DB 763 SRCVCEGPALAIIVHSOTTDIPCPHGMISLMKGFSTIMFTSAGSEGTGALAPSGCL 189
 QY 180 EEFASPLECHGRTCTNYNSYSFWLASLNPENMFRRKPIPTVKAGLEKIIISRCVC 239
 DB 823 EEFASPLECHGRTCTNYNSYSFWLASLNPENMFRRKPIPTVKAGLEKIIISRCVC 249

OY 240 MKK 242
Db 883 MKR 885

RESULT 13

O9NYC5 PRELIMINARY; PRT: 229 AA.

AC 09NYC5;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE ARRESTEN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

SEQUENCE FROM N.A.

Colorado P.C., Torre A., Kamphaus G.D., Maeshima Y., Hopfer H.,
Takahashi K., Volk R., Zamborsky E.D., Herman S., Sarkar P.K.,
Ra Eickens M.B., Dhanabal M., Simons M., Post M., Kufe D.,
Ra Weichselbaum R.R., Subhane V.P., Kalluri R.,
RT "Anti-angiogenic cues from vascular basement membrane collagen."
RL Cancer Res. 0:0-0(2000).
DR EMBL: AF258349; AAF72630.1; -;
FT NON_TER 1
SQ SEQUENCE 229 AA; 25331 MW; 9693QDC100A5C1D5 CRC64;

Query Match

Best Local Similarity 70.1%; Score 939; DB 4; Length 229;
Matches 161; Conservative 30; Mismatches 33; Indels 0; Gaps 0;

OY 19 GFVETRRHQTTAIPSCDEGTVPYSGSFLEFVQGNORAGODLTGSCIORFTMPLE 78
Db 5 GFVETRRHQTTAIPSCDEGTVPYSGSFLEFVQGNORAGODLTGSCIORFTMPLE 64
OY 79 CANNVNCNFRNDYSTWLTSPALMPNNAPTIGRALEPTISRCTVCEGPAIAVAHSOT 138
Db 65 CANNVNCNFRNDYSTWLTSPALMPNNAPTIGRALEPTISRCTVCEGPAIAVAHSOT 124
OY 139 TDIPPCPHGWSILMKGSFMTFSGSEGTGQALASPGSCLEFRASPLECHGRGTCTY 198
Db 125 TDIPPCPHGWSILMKGSFMTFSGSEGTGQALASPGSCLEFRASPLECHGRGTCTY 184
OY 199 YSNYSYFWLASLNPENRFRKPISTVAGLEKTIISRCOVCMKK 242
Db 185 YSNYSYFWLASLNPENRFRKPISTVAGLEKTIISRCOVCMKK 228

RESULT 14

O28271 PRELIMINARY; PRT: 225 AA.

AC 028271;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE COLLAGEN TYPE IV ALPHA 1 CHAIN (FRAGMENT).
GN COL4A1.

OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SAMOED;
RA MIMLINE=96278820; PubMed=8662866;
RT Thorne P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
of collagen type IV. Evidence from a canine model of X-linked
nephritis with a COL4A5 gene mutation.";

RT J. Biol. Chem. 271:13821-13828(1996).
DR EMBL: U50933; AAC48583.1; -;
DR INTERPRO: IPR001442; -;
DR PFAM: PF01413; C4; 2.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 225 AA; 24586 MW; 2196455890416E47 CRC64;

Query Match

Best Local Similarity 66.3%; Score 889; DB 6; Length 225;
Matches 155; Conservative 30; Mismatches 36; Indels 2; Gaps 1;

OY 7 GDSGSPATWTRGFVTRHSQTATPSCDEGTVPYSGSFLEFVQGNORAGODLTGSG 66
Db 1 GDSGSPATWTRGFVTRHSQTATPSCDEGTVPYSGSFLEFVQGNORAGODLTGSG 58
OY 67 CLOEFTTTPFLFCNVNDVCFASRNDYSYSLTPALMPNNAPTIGRALEPTISRCTVCE 126
Db 59 CLOEFTTTPFLFCNVNDVCFASRNDYSYSLTPALMPNNAPTIGRALEPTISRCTVCE 118
OY 127 GPALIAVAHSQTTDIPPCPHGWSILMKGSFMTFSGSEGTGQALASPGSCLEFRASP 186
Db 119 GPALIAVAHSQTTDIPPCPHGWSILMKGSFMTFSGSEGTGQALASPGSCLEFRASP 178
OY 187 FLECHGRGTCTYNSYSYFWLASLNPENRFRKPISTVAGLE 229
Db 179 FLECHGRGTCTYNSYSYFWLASLNPENRFRKPISTVAGLE 221

RESULT 15

O07265 PRELIMINARY; PRT: 1752 AA.

DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE PROCOLLAGEN ALPHA 1(III) CHAIN.
GN COL3A1ALPHA.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Euechinoidae; Echinacea; Echinoida; Strongylocentrotidae;
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93186842; PubMed=844489;
RA Exposito J.Y., D'Alessio M., Di Liberto M., Ramirez F.;
RT "Complete primary structure of a sea urchin type IV collagen alpha
chain and analysis of the 5' end of its gene."
RL J. Biol. Chem. 268:5249-5254(1993).
DR EMBL: L02917; AAA30039.1; -;
DR INTERPRO: IPR000087; -;
DR INTERPRO: IPR001442; -;
DR PFAM: PF01391; Collagen; 22.
DR PFAM: PF01413; C4; 2.
DR PRODOM: PD003923; -; 2.
KW Extracellular matrix; Collagen.
SQ SEQUENCE 1752 AA; 170210 MW; 1A55AA21569346D CRC64;

Query Match

Best Local Similarity 62.4%; Score 836; DB 5; Length 1752;
Matches 145; Conservative 37; Mismatches 59; Indels 2; Gaps 1;

OY 1 GLKGRGDSGSPATWTRGFVTRHSQTATPSCDEGTVPYSGSFLEFVQGNORAGOD 60
Db 1512 GLKGRGDSGSPATWTRGFVTRHSQTATPSCDEGTVPYSGSFLEFVQGNORAGOD 1569
OY 61 LGTGSCLQRTTTPFLFCNVNDVCFASRNDYSYSLTPALMPNNAPTIGRALEPTIS 120
Db 1570 LGTGSCLQRTTTPFLFCNVNDVCFASRNDYSYSLTPALMPNNAPTIGRALEPTIS 1629

Sat Apr 14 08:43:45 2001

us-09-543-371-10.open.rspt

Page 7

```

OY 421 RCTCEBPAIALNVHSTOTDIPRCPHGWLSJMKKGSEFIETFSAGSEBGTQALASPGSLE 180
    | | | | | : | | | | | : | | | | | : | | | | | :
Db 1630 RCVCCEAPAVULYTHVHSQVNIIPCCDPBRCWGLMGVSEFMNHTDPRGEGSGOMISFPGSLE 1689
OY 181 EFRASPLECHGRGTCNVYSNSYFWLASINPBRMPKPISTVAKGELEKISRCOVCM 240
    : | | | | | : | | | | | : | | | | | : | | | | | :
Db 1690 DFRNSPIECHGDKCNYATTYTFWLSITGNADFTMPQSETLAGSLRTRVRSRCAYCL 1749
OY 241 KKR 243
    : :
Db 1750 RNO 1752

```

Search completed: April 13, 2001, 15:45:37
Job time: 315 sec

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: April 13, 2001, 15:41:27 ; Search time 20.87 Seconds

(without alignments)
400.495 Million cell updates/sec

Title: US-09-543-371-10
Perfect score: 1340
Sequence: 1 GIKGKRGDSGSPATWTTTGRF.....KAGELPKIISRCQVCKMKRRH 244

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1340	100.0	1670	1 CA34_HUMAN	Q01955 homo sapien
2	1210.5	90.3	471	1 CA34_BOVIN	Q28084 bos taurus
3	960	71.6	1669	1 CA14_HUMAN	P02462 homo sapien
4	951	71.0	1669	1 CA14_MOUSE	P02463 mus musculus
5	943.5	70.4	1685	1 CA54_HUMAN	P29400 homo sapien
6	922.5	68.8	754	1 CA54_CANFA	Q28247 canis fami
7	845.5	63.1	1758	1 CA14_CAEEL	P17139 caenorhabdi
8	796.5	59.4	1763	1 CA24_ASCSU	P27393 ascaris suu
9	772.5	57.6	1758	1 CA24_CAEEL	P17140 caenorhabdi
10	760.5	56.8	1707	1 CA24_MOUSE	P08122 mus musculus
11	755.5	56.4	1712	1 CA24_HUMAN	P08572 homo sapien
12	739	55.1	1678	1 CA64_HUMAN	Q14031 homo sapien
13	709	45.3	453	1 CA44_BOVIN	Q29442 bos taurus
14	702	52.4	623	1 CA44_RABIT	P55787 oryctolagus
15	696	51.9	1690	1 CA44_HUMAN	P53420 homo sapien
16	694.5	51.8	1775	1 CA14_DROME	P08120 drosophila
17	98.5	7.4	537	1 COX1_SCHPO	P07657 schizosacch
18	94.5	7.1	4393	1 PGBM_HUMAN	P08166 homo sapien
19	89.5	6.7	3707	1 PGBM_MOUSE	Q05793 mus musculus
20	86.5	6.5	356	1 Y429_HUMAN	Q43312 homo sapien
21	81.5	6.1	539	1 IL2B_MOUSE	P16297 mus musculus
22	80	5.9	1597	1 GP63_LEICH	P15706 leishmania
23	79.5	5.9	1597	1 M3K4_MOUSE	Q08648 mus musculus
24	79	5.9	837	1 GCSR_MOUSE	P40223 mus musculus
25	79	5.9	1877	1 PCK7_MOUSE	Q04592 mus musculus
26	77.5	5.8	669	1 MYBE_AVILE	P01105 avian leuko
27	77.5	5.8	1328	1 AGRI_DISOM	Q05040 discopys o
28	77.5	5.8	1391	1 N155_HUMAN	Q07564 homo sapien
29	77	5.7	366	1 CAS4_ERPMU	P18503 ephyatilia m
30	77	5.7	368	1 DNAL_NITEU	O06431 nltiosomona
31	77	5.7	1639	1 LMGI_DROME	P15215 drosophila
32	76.5	5.7	155	1 HOPD_SALTY	O08927 salmonella
33	76.5	5.7	680	1 CA1A_HUMAN	Q03692 homo sapien

34	76.5	5.7	770	1 PCK7_MOUSE	O61139 mus musculus
35	76.5	5.7	783	1 PCK7_RAT	Q62849 rattus norv
36	76	5.7	379	1 DNAL_LEGPN	P50025 legionella
37	76	5.7	674	1 CALA_CHICK	P08125 gallus galli
38	75.5	5.6	785	1 PCK7_HUMAN	Q16549 homo sapien
39	75.5	5.6	1411	1 Y297_HUMAN	O15040 homo sapien
40	75.5	5.6	1959	1 AGRI_RAT	P25304 rattus norv
41	75	5.6	909	1 AD12_HUMAN	Q43184 homo sapien
42	74.5	5.6	525	1 N062_RAT	P17955 rattus norv
43	74.5	5.6	610	1 LKHA_CAVPO	P19602 cavia porce
44	74	5.5	312	1 F3ST_FLABI	P52835 flaveria bi
45	74	5.5	375	1 DNAL_ECOLI	P08622 escherichia

ALIGNMENTS

RESULT	ID	CA34_HUMAN	STANDARD	PRT	1670 AA.
AC	Q01955				
DT	01-OCT-1996	(Rel. 34, Created)			
DT	15-JUL-1999	(Rel. 38, Last sequence update)			
DT	01-OCT-2000	(Rel. 40, Last annotation update)			
DE	COLLAGEN ALPHA 3(IV) CHAIN PRECURSOR.				
GN	COL4A3				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Kidney;				
RX	MEDLINE=94364994; PubMed=8083201.				
RA	Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Reiders S.T.;				
RT	"Complete primary structure of the human alpha 3(IV) collagen chain.				
RT	Coeexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in				
RT	human tissues."				
RL	J. Biol. Chem. 269:23013-23017(1994).				
RN	[2]				
RP	REVISED.				
RA	Leinonen A.;				
RL	Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.				
RX	MEDLINE=95015826; PubMed=1400291.				
RA	Quinones S., Bernat D., Garcia-Sogo M., Elena S.F., Saus J.;				
RT	"Exon/intron structure of the human alpha 3(IV) gene encompassing the				
RT	Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially				
RT	antigenic region at the triple helix/NC1 domain junction."				
RL	J. Biol. Chem. 267:19780-19784(1992).				
RN	[4]				
RP	SEQUENCE OF 1453-1670 FROM N.A.				
RX	MEDLINE=91353570; PubMed=1882840.				
RA	Morriison K.E., Mariyama M., Yang-Feng T.L., Reiders S.T.;				
RT	"Sequence and localization of a partial cDNA encoding the human alpha				
RT	3 chain of type IV collagen."				
RL	Am. J. Hum. Genet. 49:545-554(1991).				
RN	[5]				
RP	SEQUENCE OF 1331-1670 FROM N.A.				
RC	TISSUE=Kidney;				
RX	MEDLINE=92147878; PubMed=1737849;				
RA	Turner N., Mason P.C., Brown R., Fox M., Povey S., Rees A.,				
RA	Pusey C.D.;				
RT	"Molecular cloning of the human Goodpasture antigen demonstrates it				
RT	to be the alpha 3 chain of type IV collagen."				
RL	J. Clin. Invest. 89:592-601(1992).				
RN	[6]				
RP	SEQUENCE OF 1644-1670 FROM N.A.				
RC	TISSUE=Kidney;				
RA	Ding J.;				
RL	Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.				
RN	[7]				

RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.
RC TISSUE-Kidney;
RX MEDLINE=94124597; PubMed=8294492;
RA Feng L., Xia Y., Wilson C.B.,
RT "Alternative splicing of the NC1 domain of the human alpha 3(IV)
RT collagen gene. Differential expression of mRNA transcripts that
RT predict three protein variants with distinct carboxyl regions";
RL J. Biol. Chem. 269:2342-2348(1994).
RN [8]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=98196854; PubMed=9537506;
RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,
RA Niinomiya Y.,
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome
RT 2q36";
RL FEBS Lett. 424:11-16(1998).
RN [9]
RP ALTERNATIVE SPLICING.
RX MEDLINE=93280184; PubMed=8505332;
RA Bernal D., Quiñones S., Saus J.;
RA "The human mRNA encoding the Goodpasture antigen is alternatively
RT spliced";
RL J. Biol. Chem. 268:12090-12094(1993).
RN [10]
RP VARIANT PRO-1474.
RX MEDLINE=95078827; PubMed=7987301;
RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,
RA Barrientos A., Monnens L.A.H., van Oost B.A., Brunner H.G.,
RA Reinders S.T., Smeets H.J.M.,
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal
RT recessive Alport syndrome";
RL Hum. Mol. Genet. 3:1269-1273(1994).
CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
CC NIDOCEN.
CC
CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS. ALPHA 1(IV)-
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC
CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS: 1 (SHOWN HERE), 2/V AND
CC 3/LS. ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR
CC C-TERMINAL NC1 DOMAINS.
CC
CC -1- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
CC COLLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
CC COCHLEA, LUNG AND BRAIN.
CC
CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC
CC -1- PPM: PROLINES AT THE THIRD POSITION OF THE TRIPLEPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC
CC -1- PPM: THE ALTERNATIVE SPLICED FORM V CONTAINS AN ADDITIONAL
CC N-LINKED GLYCOSYLATION SITE.
CC
CC -1- PPM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTR- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC
CC -1- DISEASE: ANTIBODIES AGAINST THE NC1 DOMAIN OF ALPHA3(IV) MEDIATE
CC THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS
CC CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.
CC
CC -1- DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE I
CC AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY
CC GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,
CC HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN
CC MALES AND FEMALES.
CC
CC -1- THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way

CC		modified and this statement is not removed.	Usage by and for commercial
CC		entities requires a license agreement (see	http://www.isb-slb.ch/announce/
CC		or send an email to:	license@isb-slb.ch).
CC			
DR	EMBL; X80031; CAAS6335.1;	-	
DR	EMBL; M92993; AAA21610.1;	-	
DR	EMBL; S55790; AAB19637.1;	-	
DR	EMBL; M81379; AAAS1556.1;	-	
DR	EMBL; L08650; AAAS2044.1;	-	
DR	EMBL; U02519; AAA18942.1;	-	
DR	EMBL; U02520; AAA18943.1;	-	
DR	EMBL; AB008495; BAA25064.1;	-	
DR	MIM; 120070;	-	
DR	MIM; 203780;	-	
DR	MIM; 233450;	-	
DR	InterPro; IPR000087;	-	
DR	InterPro; IPR001442;	-	
DR	Pfam; PF01413; C4; 2.		
DR	Pfam; PF01391; Collagen; 21.		
KW	Extracellular matrix; Connective tissue; Repeat; Hydroxylation;		
KV	Glycoprotein; Basement membrane; Collagen; Signal; Cell adhesion;		
KW	Alternative splicing; Polymorphism; Phosphorylation.		
FT	SIGNAL	1	28
FT	CHAIN	29	1670
FT	DOMAIN	29	42
FT	DOMAIN	43	1438
FT	DOMAIN	1439	1670
FT	DOMAIN	1427	1444
FT	SITE	1426	1427
FT	SITE	791	793
FT	SITE	996	998
FT	SITE	1154	1156
FT	SITE	1305	1308
FT	SITE	1345	1347
FT	SITE	1432	1434
FT	CARBOHYD	253	253
FT	MOD_RES	1435	1435
FT	MOD_RES	1437	1437
FT	DISULFD	1460	1551
FT	DISULFD	1493	1548
FT	DISULFD	1505	1511
FT	DISULFD	1570	1665
FT	DISULFD	1604	1662
FT	DISULFD	1616	1622
FT	VARSPLIC	1586	1670
FT	VARSPLIC	1488	1670
FT	VARIANT	1474	1474
FT	VARIANT	1495	1495
FT	CONFLICT	1539	1539
FT	CONFLICT	1594	1594
FT	CONFLICT	1663	1664
SO	SEQUENCE	1670 AA; 161740 MM; 9B95700AF399BEAE CRC64;	
QY	Query Match	100.0%; Score 1340; DB 1; Length 1670;	
	Best Local Similarity	100.0%; Pred. No. 1.3e-118;	
	Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
	1 GLKGRGDSGFATWTTRGFTFRHSQTATAISCPBGIVPLYSGSFELVGCGNORAHGD 60		

```

b:
DB 1427 GLKGRGSDSGSPATWTT-RCGFVTRHSGTQTAIPSCPECTVLYSGSFLEFVQGNORAHGQD 1486
QY 61 LGTLGSCLOFRTMPFLFCNVNDVNCNFSANDYSYWLSTPALMPMNAPIPTGRALPEYIS 120
DB 1487 LGTLGSCLOFRTMPFLFCNVNDVNCNFSANDYSYWLSTPALMPMNAPIPTGRALPEYIS 1546
QY 121 RCIVCEGPALAIIVHSGTQTDIPPCPHGMISLWKGFSIFMTSAGSEGTGALASPGSCLE 180
DB 1547 RCIVCEGPALAIIVHSGTQTDIPPCPHGMISLWKGFSIFMTSAGSEGTGALASPGSCLE 1606
QY 181 EFRASPFLECHGRGTCNYNSYSFWMLASLNPERMFRKPIPTVKAGELKTIISRCQVCM 240
DB 1607 EFRASPFLECHGRGTCNYNSYSFWMLASLNPERMFRKPIPTVKAGELKTIISRCQVCM 1666
QY 241 KKRH 244
DB 1667 KKRH 1670

RT 2
C BOVIN STANDARD; PRT; 471 AA.
ID CA34 BOVIN
AC Q28084;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE COLLAGEN ALPHA 3(IV) CHAIN (FRAGMENT).
GN COLA43.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lens;
RX MEDLINE=91093146; PubMed=1985905;
RA Morrison K.E., Germino G.G., Reeders S.T.;
RT "Use of the polymerase chain reaction to clone and sequence a cDNA
RT encoding the bovine alpha 3 chain of type IV collagen.";
RL J. Biol. Chem. 266:34-39(1991).
RN [2]
RP SEQUENCE OF 227-258.
RC TISSUE=Kidney;
RX MEDLINE=90202779; PubMed=2318822;
RA Gunner S., Saus J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
RT alpha 4, of type IV collagen.";
RL J. Biol. Chem. 265:5466-5469(1990).
RN [3]
RP SEQUENCE OF 227-254.
RX MEDLINE=88330844; PubMed=3417661;
RA Saus J., Wieselander J., Langeveld J.P.M., Quinones S., Hudson B.G.;
RT "Identification of the Goodpasture antigen as the alpha 3(IV) chain
RT of collagen IV.";
RL J. Biol. Chem. 263:13374-13380(1988).
RN [4]
RP SEQUENCE OF 227-244.
RX MEDLINE=87222419; PubMed=2438283;
RA Butkowsk R.J., Langeveld J.P.M., Wieselander J., Hamilton J.,
RA Hudson B.G.;
RT "Localisation of the Goodpasture epitope to a novel chain of basement
RT membrane collagen.";
RL J. Biol. Chem. 262:7874-7877(1987).
CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
CC NIDOGEN.
CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS

```

```

CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTRA- AND INTERMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M63139; AAA62708.1; -.
DR InterPro: IPR000087; -.
DR InterPro: IPR001442; -.
DR Pfam: PF01413; C4; 2.
DR Pfam: PF01391; Collagen: 4.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Cell adhesion.
FT NON_TER 1
FT DOMAIN <1 238 TRIPLE-HELICAL REGION.
FT SITE 239 471 NONHELICAL REGION (NC1).
FT MOD_RES 106 108 CELL ATTACHMENT SITE (POTENTIAL).
FT MOD_RES 232 232 HYDROXYLATION.
FT MOD_RES 238 238 HYDROXYLATION.
FT DISULFID 261 352 OR 349 (BY SIMILARITY).
FT DISULFID 294 349 OR 352 (BY SIMILARITY).
FT DISULFID 306 312 BY SIMILARITY.
FT DISULFID 371 466 OR 463 (BY SIMILARITY).
FT DISULFID 405 463 OR 466 (BY SIMILARITY).
FT DISULFID 417 423 BY SIMILARITY.
FT CONFLICT 253 253 S -> Y (IN REF. 3).
SQ SEQUENCE 471 AA; 47585 MW; C03B6F14E7008DE CRC64;

Query Match 90.3%; Score 1210.5; DB 1; Length 471;
Best Local Similarity 90.6%; Pred. No. 5.4e-107;
Matches 221; Conservative 10; Mismatches 12; Indels 1; Gaps 1;

QY 1 GLKGRGSDSGSPATWTT-RCGFVTRHSGTQTAIPSCPECTVLYSGSFLEFVQGNORAHGQ 59
DB 227 GLKGRGSDTGPAGAVWGRGFVTRHSGTQTAIPSCPECTVLYSGSFLEFVQGNORAHGQ 286
QY 60 DLGTLGSCLOFRTMPFLFCNVNDVNCNFSANDYSYWLSTPALMPMNAPIPTGRALPEYI 119
DB 287 DLGTLGSCLOFRTMPFLFCNVNDVNCNFSANDYSYWLSTPALMPMNAPIPTGRALPEYI 346
QY 120 SRCIVCEGPALAIIVHSGTQTDIPPCPHGMISLWKGFSIFMTSAGSEGTGALASPGSCLE 179
DB 347 SRCIVCEGPALAIIVHSGTQTDIPPCPHGMISLWKGFSIFMTSAGSEGTGALASPGSCLE 406
QY 180 EFRASPFLECHGRGTCNYNSYSFWMLASLNPERMFRKPIPTVKAGELKTIISRCQVCM 239
DB 407 EFRASPFLECHGRGTCNYNSYSFWMLASLNPERMFRKPIPTVKAGELKTIISRCQVCM 466
QY 240 KKRH 243
DB 467 KKRH 470

RESULT 3
CA14_HUMAN STANDARD; PRT; 1669 AA.
ID CA14_HUMAN
AC P02462;
DT 21-JUL-1986 (Rel. 01, Created)

```

DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DE 01-OCT-2000 (Rel. 40, Last annotation update)
 DE COLLAGEN ALPHA 1(IV) CHAIN PRECURSOR.
 GN COL4A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCBI_TaxID=9606;
 RP [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89340433; PubMed=2701944;
 RA Solinen R., Huotari M., Ganguly A., Prockop D.J., Tryggvason K.;
 RT "Structural organization of the gene for the alpha 1 chain of human
 RT type IV collagen.";
 RL J. Biol. Chem. 264:13565-13571(1989).
 RN [2]
 RN SEQUENCE OF 46-1257 FROM N.A.
 RP TISSUE=Placenta;
 RC MEDLINE=88083584; PubMed=3691802;
 RX Solinen R., Haka-Risku T., Prockop D.J., Tryggvason K.;
 RA "Complete primary structure of the alpha 1-chain of human basement
 RA membrane (type IV) collagen.";
 RL FEBS Lett. 225:188-194(1987).
 RN [3]
 RN SEQUENCE OF 1-943 FROM N.A.
 RP TISSUE=Placenta;
 RC MEDLINE=88029471; PubMed=3311751;
 RX Brazel D., Oberhauser I., Dierlinger H., Babel W., Glanville R.W.,
 RA Deutzmann R., Kuehn K.;
 RT "Completion of the amino acid sequence of the alpha 1 chain of human
 RT basement membrane collagen (type IV) reveals 21 non-triplet
 RT interruptions located within the collagenous domain.";
 RL Eur. J. Biochem. 168:529-536(1987).
 RN [4]
 RN SEQUENCE OF 28-243.
 RP MEDLINE=86004708; PubMed=4043082;
 RX Glanville R.W., Qian R.Q., Siebold B., Risteli J., Kuehn K.;
 RA "Amino acid sequence of the N-terminal aggregation and cross-linking
 RA region (7S domain) of the alpha 1 (IV) chain of human basement
 RA membrane collagen.";
 RL Eur. J. Biochem. 152:213-219(1985).
 RN [5]
 RN SEQUENCE OF 534-1447.
 RP MEDLINE=8503629; PubMed=6434307;
 RX Babel W., Glanville R.W.;
 RA "Structure of human basement-membrane (type IV) collagen. Complete
 RA amino-acid sequence of a 914-residue-long pepsin fragment from the
 RA alpha 1(IV) chain.";
 RL Eur. J. Biochem. 143:545-556(1984).
 RN [6]
 RN SEQUENCE OF 1256-1669 FROM N.A.
 RP MEDLINE=85207819; PubMed=2581969;
 RX Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R.,
 RA Cheung M.-C., Prockop D.J., Boyd C.D.;
 RT "cDNA clones coding for the pro-alpha1(IV) chain of human type IV
 RT procollagen reveal an unusual homology of amino acid sequences in two
 RT halves of the carboxyl-terminal domain.";
 RL J. Biol. Chem. 260:7681-7687(1985).
 RN [7]
 RN SEQUENCE OF 1259-1669 FROM N.A.
 RP MEDLINE=85216555; PubMed=2582422;
 RX Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenbloom J.,
 RA Kefauver N.A., Myers J.C.;
 RT "Restricted homology between human alpha 1 type IV and other
 RT procollagen chains.";
 RL Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).
 RN [8]
 RN SEQUENCE OF 1-28 FROM N.A.
 RP MEDLINE=89034231; PubMed=3182844;
 RX Solinen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
 RA "The structural genes for alpha 1 and alpha 2 chains of human type IV
 RA collagen are divergently encoded on opposite DNA strands and have an
 RT overlapping promoter region.";

RL J. Biol. Chem. 263:17217-17220(1988).
 RN [9]
 RN SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
 RP TISSUE=Placenta;
 RC MEDLINE=89005112; PubMed=2844531;
 RX Siebold B., Deutzmann R., Kuehn K.;
 RT "The arrangement of intra- and intermolecular disulfide bonds in the
 RT carboxyterminal, non-collagenous aggregation and cross-linking domain
 RT of basement-membrane type IV collagen.";
 RL Eur. J. Biochem. 176:617-624(1988).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 CC NIDOGEN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOCFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL, TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: LYSINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT ARE HYDROXYLATED IN ALL CASES AND BIND CARBOHYDRATE.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M26576; AAA53098.1; JOINED.
 DR EMBL; J04217; AAA53098.1; JOINED.
 DR EMBL; M26550; AAA53098.1; JOINED.
 DR EMBL; M26540; AAA53098.1; JOINED.
 DR EMBL; M26542; AAA53098.1; JOINED.
 DR EMBL; M26543; AAA53098.1; JOINED.
 DR EMBL; M26544; AAA53098.1; JOINED.
 DR EMBL; M26545; AAA53098.1; JOINED.
 DR EMBL; M26546; AAA53098.1; JOINED.
 DR EMBL; M26547; AAA53098.1; JOINED.
 DR EMBL; M26537; AAA53098.1; JOINED.
 DR EMBL; M26538; AAA53098.1; JOINED.
 DR EMBL; M26548; AAA53098.1; JOINED.
 DR EMBL; M26549; AAA53098.1; JOINED.
 DR EMBL; M26551; AAA53098.1; JOINED.
 DR EMBL; M26552; AAA53098.1; JOINED.
 DR EMBL; M26553; AAA53098.1; JOINED.
 DR EMBL; M26554; AAA53098.1; JOINED.
 DR EMBL; M26555; AAA53098.1; JOINED.
 DR EMBL; M26556; AAA53098.1; JOINED.
 DR EMBL; M26557; AAA53098.1; JOINED.
 DR EMBL; M26558; AAA53098.1; JOINED.
 DR EMBL; M26559; AAA53098.1; JOINED.
 DR EMBL; M26560; AAA53098.1; JOINED.
 DR EMBL; M26561; AAA53098.1; JOINED.
 DR EMBL; M26562; AAA53098.1; JOINED.
 DR EMBL; M26563; AAA53098.1; JOINED.
 DR EMBL; M26564; AAA53098.1; JOINED.
 DR EMBL; M26565; AAA53098.1; JOINED.
 DR EMBL; M26566; AAA53098.1; JOINED.
 DR EMBL; M26567; AAA53098.1; JOINED.

44

```

DR EMBL; M26568; AAA53098.1; JOINED.
DR EMBL; M26569; AAA53098.1; JOINED.
DR EMBL; M26570; AAA53098.1; JOINED.
DR EMBL; M26571; AAA53098.1; JOINED.
DR EMBL; M26572; AAA53098.1; JOINED.
DR EMBL; M26573; AAA53098.1; JOINED.
DR EMBL; M26574; AAA53098.1; JOINED.
DR EMBL; M26575; AAA53098.1; JOINED.
DR EMBL; Y00706; CAA68698.1; JOINED.
DR EMBL; X05561; CAA68698.1; JOINED.
DR EMBL; M10940; AAA52006.1; JOINED.
DR EMBL; M11315; AAA52042.1; JOINED.
DR PIR; S16876; CGH04B.
DR MIM; 120130.
DR InterPro; IPR000087.
DR InterPro; IPR001442.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 21.
KW Extracellular matrix; Connective tissue; Basement membrane;
K Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
FT SIGNAL 1 27
FT PROPEP 28 172
FT CHAIN 173 1669
FT DOMAIN 173 1440
FT DOMAIN 1441 1669
FT CARBOHYD 126 126
FT DISULFID 1460 1551
FT DISULFID 1493 1548
FT DISULFID 1493 1548
FT DISULFID 1505 1511
FT DISULFID 1570 1665
FT DISULFID 1604 1662
FT DISULFID 1616 1622
FT CONFLICT 237 238
FT CONFLICT 241 249
FT CONFLICT 319 319
FT CONFLICT 719 719
FT CONFLICT 837 837
FT CONFLICT 842 842
FT CONFLICT 896 896
FT CONFLICT 904 904
FT CONFLICT 914 914
FT CONFLICT 998 998
FT CONFLICT 1010 1010
FT CONFLICT 1012 1012
FT CONFLICT 1358 1358
SQ SEQUENCE 1669 AA; 160611 MW; 3BEBADFFB9B8A84 CRC64;

SG -> KE (IN REF. 4).
G -> A (IN REF. 4).
Q -> A (IN REF. 3).
N -> D (IN REF. 5).
D -> Y (IN REF. 5).
K -> P (IN REF. 5).
V -> W (IN REF. 2).
E -> Q (IN REF. 5).
S -> K (IN REF. 5).
K -> P (IN REF. 5).
S -> K (IN REF. 5).
E -> Q (IN REF. 5).
OR 1548.
OR 1551.
OR 1662.
OR 1665.

```

Very Match 71.6%; Score 960; DB 1; Length 1669;
 est local similarity 69.0%; Pred. No. 1e-82;
 Matches 167; Conservative 32; Mismatches 41; Indels 2; Gaps 1;

```

QY 1 GLKGRKDGSPATWTRGFTVRHSQTALIPSCPECTVLYSGFSLFVQGNQRAHGOD 60
DB 1429 GLGSGMGPPTPS--VDHGFVTRHSQTDPOCPSCGTKLHYGXYLLYQGNBRAGOD 1486
QY 61 LGLGSLQFTMPPLFCVNDVNCNFRANDYSYMLSTALMNMNAPITGRLEYIS 120
DB 1487 LGRAGSLRFRSTMPPLFCVNDVNCNFRANDYSYMLSTALMNMNAPITGRLEYIS 1546
QY 121 RCVCSEPAFAIVHSOTITPPCPHGMISLWKGFSFIMTSGASEGSGOALASPGSCLE 180
DB 1547 RCACVCEPAMVAVHSOTITPPCPHGMISLWKGFSFIMTSGASEGSGOALASPGSCLE 1606
QY 181 EFRASPLECHGSGTCTMYNSYSYFWLASLNPBERFKRPISPVKAGLEKTIISRCQVCM 240
DB 1607 EFRASPLECHGSGTCTMYNSYSYFWLASLNPBERFKRPISPVKAGLEKTIISRCQVCM 1666
QY 241 KK 242
DB 1667 RR 1668

```

```

RESULT 4
CAL_MOUSE
ID CAL1_MOUSE STANDARD; PRT; 1669 AA.
AC P02463;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE COLLAGEN ALPHA 1(IV) CHAIN PRECURSOR.
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197932; PubMed=2703490.
RA Muthukumar G., Blumberg B., Kurkunen M.;
RT "The complete primary structure for the alpha 1-chain of mouse
RT collagen IV. Differential evolution of collagen IV domains.";
RL J. Biol. Chem. 264:6310-6317(1989).
RN [2]
RP SEQUENCE OF 1-1154 FROM N.A.
RX MEDLINE=88112221; PubMed=3338568;
RA Wood L., Theriault N., Vogeli G.;
RT "cDNA clones complementing the nucleotide and derived amino acid
RT sequence of the alpha 1 chain of basement membrane (type IV) collagen
RT from mouse.";
RL FEBS Lett. 227:5-8(1988).
RN [3]
RP SEQUENCE OF 1149-1424 FROM N.A.
RX MEDLINE=86301886; PubMed=3755692;
RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
RT synthetic oligodeoxynucleotide.";
RN [4]
RP SEQUENCE OF 1276-1669 FROM N.A.
RX MEDLINE=85127033; PubMed=2578661;
RA Oberbauer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
RA Vogeli G., Voss T., Stebold B., Glanville R.W., Kuhn K.;
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
RT the alpha 1(IV) chain of basement membrane collagen as derived from
RT complementary DNA.";
RL Eur. J. Biochem. 147:217-224(1985).
RN [5]
RP SEQUENCE OF 1441-1669 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkunen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86196099; PubMed=3009468;
RA Sakurai Y., Sullivan M., Yamada Y.;
RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
RT collagen genes.";
RL J. Biol. Chem. 261:6654-6657(1986).
RN [7]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kayes P., Wood L., Theriault N., Kurkunen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burdeto P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
RN [9]

```

RP SEQUENCE OF 1-129 FROM N.A.
 RX MEDLINE=88243724; PubMed=3379041;
 RA Kallen P.D., Burbelo P., Sakurai Y., Yamada Y.;
 RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
 RL collagen chain and the corresponding region of the gene";
 J. Biol. Chem. 263:8706-8709(1988).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 CC NIDOGEN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PMW: PROLINES AT THE THIRD POSITION OF THE TRIPLEPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: J03758; AAA37439.1; -;
 DR EMBL: M23333; AAA51625.1; -;
 DR EMBL: J04694; AAA50292.1; -;
 DR EMBL: X06777; CAA29946.1; -;
 DR EMBL: X02201; CAA26132.1; -;
 DR EMBL: M15832; AAA37340.1; -;
 DR EMBL: M14042; AAA37342.1; -;
 DR EMBL: M12879; AAA37343.1; -;
 DR EMBL: M13024; -; NOT_ANNOTATED_CDS.
 DR EMBL: M13025; -; NOT_ANNOTATED_CDS.
 DR EMBL: M13026; AAA37344.1; -;
 DR EMBL: M13027; AAA37345.1; -;
 DR EMBL: M13043; AAA37346.1; -;
 DR EMBL: J04448; AAA37437.1; -;
 DR PIR: A33525; CGMS4B.
 DR MG: MGI:88454; COL4A1.
 DR InterPro: IPR000087; -;
 DR InterPro: IPR001442; -;
 DR Pfam: PF01413; C4; 2.
 DR Pfam: PF01391; C4; 2.
 KW Extracellular matrix; Connective tissue; Basement membrane;
 KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 27
 FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
 FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.
 FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.
 FT 1441 1669 NONHELICAL REGION (NC1).
 FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).
 FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).
 FT DISULFID 1505 1511 BY SIMILARITY.
 FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).
 FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).
 FT DISULFID 1616 1622 BY SIMILARITY.
 FT CARBOHYD 126 126 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT 26 26 A -> P (IN REF. 2).
 FT 186 186 S -> L (IN REF. 2).
 FT 319 319 Q -> S (IN REF. 2).
 FT 369 369 Q -> L (IN REF. 2).
 FT 403 403 L -> F (IN REF. 2).
 FT 481 481 P -> L (IN REF. 2).
 FT CONFLICT 481 481

FT CONFLICT 493 493 Q -> H (IN REF. 2).
 FT CONFLICT 712 712 S -> I (IN REF. 2).
 FT CONFLICT 813 813 E -> O (IN REF. 2).
 FT CONFLICT 982 982 Q -> H (IN REF. 2).
 FT CONFLICT 1397 1397 V -> S (IN REF. 3).
 SQ SEQUENCE 1669 AA; 160680 MW; 42916B91E52058BE9 CRC64;
 Query Match 71.0%; Score 951; DB 1; Length 1669;
 Best Local Similarity 68.6%; Pred. No. 7.3e-82;
 Matches 166; Conservative 34; Mismatches 40; Indels 2; Gaps 1;
 QY 1 GLGKRGKDGSPATWTTTRFEVFRHSQTALPSCPAGTPLYSGFSEFLPVGNORAGHD 60
 DB 1429 GLPGSGPPGTPS--VDHGLVTRHSQTDDPLCPGCTKLILGYSLIVQGERAHGD 1486
 QY 61 LGTIGSCLOFTTTPPLFCNVNDVNCNPASRNDYSYLSLTPALPMNMAPITGALPEYIS 120
 DB 1487 LGTAGSCLEKFTSTMPPLFCNNVNCNMFASRNDYSYLSLTPPEPMGMAPISGDNIRPIS 1546
 QY 121 RCTVCGPAILAIVHSGOTDIPPCPHGWTSLMKGFSEFIMFTSGSGTGOALSPGSCLE 180
 DB 1547 RCAVCCAPAVMAVHSGOTDIPCCPMGWSLMIGSEFVHTTSAGAGSGOALSPGSCLE 1606
 QY 181 EFRASPFLCHGRGTCNYSNSYSFWLASLNPBRMRKPIPTVKADELKTIISRCQVCN 240
 DB 1607 EFRSAPFLCHGRGTCNYSNSYSFWLASLNPBRMRKPIPTVKADELKTIISRCQVCN 1666
 QY 241 KK 242
 DB 1667 RR 1668
 RESULT 5
 CAS4_HUMAN STANDARD; PRT; 1685 AA.
 ID CAS4_HUMAN
 AC P29400; Q16126; Q16006;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE COLLAGEN ALPHA 5(IV) CHAIN PRECURSOR.
 GN COL4A5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94165049; PubMed=8120014;
 RA Zhou J., Leinonen A., Tryggvason K.;
 RT "Structure of the human type IV collagen COL4A5 gene";
 RL J. Biol. Chem. 269:6608-6614(1994).
 RN [2]
 RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.
 RX TISSUE=Kidney;
 RX MEDLINE=92316923; PubMed=1352287;
 RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
 RT "Complete amino acid sequence of the human alpha 5 (IV) collagen
 chain and identification of a single-base mutation in exon 23
 converting glycine 521 in the collagenous domain to cysteine in an
 Alport syndrome patient";
 RL J. Biol. Chem. 267:12475-12481(1992).
 RN [3]
 RP SEQUENCE OF 85-1685 FROM N.A.
 RX TISSUE=Placenta;
 RX MEDLINE=90337990; PubMed=2380186;
 RA Pihlajaniemi T., Pihlajaniemi E.R., Myers J.C.;
 RT "Complete primary structure of the triple-helical region and the
 carboxyl-terminal domain of a new type IV collagen chain, alpha
 5(IV)";
 RL J. Biol. Chem. 265:13758-13766(1990).
 RN [4]
 RP SEQUENCE OF 924-1685 FROM N.A.

RX MEDLINE-91169491; PubMed-2004755;
RA Zhou J., Hostalka S.L., Chow L.T., Tryggvason K.;
RT "characterization of the 3' half of the human type IV collagen alpha
RL 5 gene that is affected in the Alport syndrome.";
RN Genomics 9:1-9(1991).
[5]
RP SEQUENCE OF 914-1685 FROM N.A.
RX MEDLINE-90160375; PubMed-1689491;
RA Hostalka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B.,
RT Tryggvason K.;
RT "Identification of a distinct type IV collagen alpha chain with
RT restricted kidney distribution and assignment of its gene to the
RT locus of X chromosome-linked Alport syndrome.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
[6]
RP SEQUENCE OF 1442-1471 FROM N.A.
RX MEDLINE-90252791; PubMed-233699;
RA Myers J.C., Jones T.A., Pohjalainen E.R., Kadri A.S., Goddard A.D.,
RT Shear D., Solomon E., Pihlajaniemi T.;
RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene
RT to the region of the X chromosome containing the Alport syndrome
RT locus.";
RL Am. J. Hum. Genet. 46:1024-1033(1990).
[7]
RP SEQUENCE OF 1-20 FROM N.A.
RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.,
RT Marynen P.;
RT Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
[8]
RP SEQUENCE OF 1258-1270 FROM N.A. (SPliced FORM).
RX MEDLINE-94133540; PubMed-8301933;
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,
RT Cassiman J., Marynen P.;
RT "Differential splicing of COL4A5 mRNA in kidney and white blood
RT cells: a complex mutation in the COL4A5 gene of an Alport patient
RT deletes the NC1 domain.";
RL Kidney Int. 44:1316-1321(1993).
[9]
RP REVIEW ON VARIANTS.
RX MEDLINE-97338662; PubMed-9195222;
RA Lemmink H.H., Schroeder C.H., Komenens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499(1997).
[10]
RP VARIANT AS SER-1564.
RX MEDLINE-91169492; PubMed-1672282;
RA Zhou J., Barker D.F., Hostalka S.L., Gregory M.C., Atkin C.L.,
RT Tryggvason K.;
RT "Single base mutation in alpha 5(IV) collagen chain converting a
RT conserved cysteine to serine in Alport syndrome.";
RN Genomics 9:10-18(1991).
[11]
RP VARIANT AS ARG-325.
RX MEDLINE-92303559; PubMed-1376965;
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Grunfeld J.-P.,
RT Tryggvason K., Gubler M.-C., Antignac C.;
RT "Substitution of arginine for glycine 325 in the collagen alpha 5
RT (IV) chain associated with X-linked Alport syndrome: characterization
RT of the mutation by direct sequencing of PCR-amplified lymphoblast
RT DNA fragments.";
RL Am. J. Hum. Genet. 51:135-142(1992).
[12]
RP VARIANT AS GLU-325.
RX MEDLINE-93244772; PubMed-1363780;
RA Renieri A., Serri M., Myers J.C., Pihlajaniemi T., Massella L.,
RT Rizzoni G., de Marchi M.;
RT "De novo mutation in the COL4A5 gene converting glycine 325 to
RT glutamic acid in Alport syndrome.";
RN Hum. Mol. Genet. 1:127-129(1992).
[13]
RP VARIANTS AS THR-1517, SER-1538 AND GLN-1563.
RX MEDLINE-94010948; PubMed-8406498;
RA Lemmink H.H., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,

RA Tryggvason K., Haggma-Schouten W.A.G., Roodvoets A.P., Rascher W.,
RA van Oost B.A., Smeets H.J.M.;
RT "Identification of four novel mutations in the COL4A5 gene of
RT patients with Alport syndrome.";
RL Genomics 17:485-489(1993).
[14]
RP VARIANTS AS E-400; V-406; V-638; R-653; R-796; R-869; R-872 & C-1241.
RX MEDLINE-95322976; PubMed-7596331;
RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;
RT "Detection of 12 novel mutations in the collagenous domain of the
RT COL4A5 gene in Alport syndrome patients.";
RL Hum. Mutat. 5:197-204(1995).
[15]
RP VARIANT AS ARG-1649.
RX MEDLINE-96213750; PubMed-8651292;
RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
RA Denison J.C., Fain P.R., Gregory M.C.;
RT "A mutation causing Alport syndrome with tardive hearing loss is
RT common in the western United States.";
RL Am. J. Hum. Genet. 58:1157-1165(1996).
[16]
RP VARIANTS AS.
RX MEDLINE-96213754; PubMed-8651296;
RA Renieri A., Bruttini M., Galli L., Zaneli P., Neri T., Rossetti S.,
RA Turco A., Hetskeri N., Zhou J., Gusmano R., Massella L., Banfi G.,
RA Scolari F., Sessa A., Rizzoni G., Tryggvason K., Pignatelli P.F.,
RA Savi M., Ballabio A., de Marchi M.;
RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
RT exons of the COL4A5 gene.";
RL Am. J. Hum. Genet. 58:1192-1204(1996).
[17]
RP VARIANT AS ASP-1498.
RX MEDLINE-96233932; PubMed-8829632;
RA Tierskaya S., Bodrynina V., Tsalykova F., Ignatova M.,
RA Krasnopolskaya X., Evgratov O.;
RT "Substitution of A1498D in noncollagen domain of $\alpha 5(IV)$ collagen
RT chain associated with adult-onset X-linked Alport syndrome.";
RL Hum. Mutat. 7:149-150(1996).
[18]
RP VARIANT AS GLN-1677.
RX MEDLINE-97295089; PubMed-9150741;
RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
RT "Common ancestry of three Ashkenazi-American families with Alport
RT syndrome and COL4A5 R1677Q.";
RL Hum. Genet. 99:681-684(1997).
[19]
RP VARIANTS AS R-174; R-177; R-325; C-1410; W-1421; T-1517 AND D-1596.
RX MEDLINE-98112435; PubMed-9452056;
RA Neri T.M., Zaneli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
RA Pignatelli G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
RA Trivelli A., Pinciaroli A.R., Ragalolo M., Rizzoni G.F., de Marchi M.;
RT "Missense mutations in the COL4A5 gene in patients with X-linked
RT Alport syndrome.";
RL Hum. Mutat. Suppl. 1:S106-S109(1998).
[20]
RP VARIANTS AS AND VARIANTS.
RX MEDLINE-99140256; PubMed-10094548;
RA Plant K.E., Green P.W., Vetric D., Flinter F.A.;
RT "Detection of mutations in COL4A5 in patients with Alport syndrome.";
RL Hum. Mutat. 13:124-132(1999).
[21]
RP FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
CC NIDOXEN.
CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS. ALPHA 1(IV)-
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2: ARE
CC PROBABLY PRODUCED BY ALTERNATIVE SPLICING. ISOFORM 2 IS A LONGER
CC ISOFORM FOUND IN KIDNEY, IN WHICH 2 EXTRA G-X-X REPEATS INTO THE
CC TRIPLE-HELIX DOMAIN ARE INTRODUCED.
CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE

SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).

1-MAR-1992 (Rel. 21, Last sequence update)

01-NOV-1997 (Rel. 35, last annotation update)
 DE COLLAGEN ALPHA 1(IV) CHAIN PRECURSOR.
 GN EMB-9 OR CIB-2 OR K04H.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 ON NCBI_TaxID=6239;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=91141582; PubMed=1996137;
 RA Guo X., Johnson J., Kramer J.M.;
 RT "Embryonic lethality caused by mutations in basement membrane
 collagen of C. elegans."
 RL Nature 349:707-709(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Milson R., Alnsough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
 RA Fulton L., Gardier A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
 RA Lettelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showken R.,
 RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulston J., Therry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Watson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans."
 RL Nature 368:32-38(1994).
 RN [3]
 RP SEQUENCE OF 1446-1758 FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=90080929; PubMed=2793871;
 RA Guo X., Kramer J.M.;
 RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
 genes are located on separate chromosomes."
 RL J. Biol. Chem. 264:17574-17582(1989).
 CC -1- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.
 CC -1- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.
 CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH
 CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1
 CC DOMAINS.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PPM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PPM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -1- DISEASE: MUTATIONS IN THIS GENE CAUSE TEMPERATURE-SENSITIVE
 CC LETHALITY DURING LATE EMBRYOGENESIS.
 CC LETHALITY DURING LATE EMBRYOGENESIS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X56979; CAA40299.1; -
 CC EMBL: Z27078; CAA81584.1; -
 CC EMBL: J05067; AAB59179.1; -
 CC DR PIR: B34476; B34476.
 CC PIR: S13651; S13651.

DR Wormpep; K04H.1; CE00246.
 DR InterPro; IPR000087; -
 DR InterPro; IPR001442; -
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 22.
 DR Extracellular matrix; Connective tissue; Basement membrane;
 KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
 FT SIGNAL 1
 FT PROPEP 2194
 FT CHAIN 2195
 FT DOMAIN 195
 FT DOMAIN 1530
 FT DISULFID 1549
 FT DISULFID 1582
 FT DISULFID 1594
 FT DISULFID 1659
 FT DISULFID 1693
 FT DISULFID 1705
 FT VARIANT 402
 FT VARIANT 408
 FT CONFLICT 1514
 FT CONFLICT 130
 FT CONFLICT 259
 FT CONFLICT 302
 FT CONFLICT 366
 FT CONFLICT 581
 FT CONFLICT 768
 FT CONFLICT 813
 FT CONFLICT 830
 FT CONFLICT 1275
 FT CONFLICT 1275
 FT CONFLICT 1722
 FT CONFLICT 1722
 SQ SEQUENCE 1758 AA; 170857 MW; 7083D9AF63E05D45 CRC64;
 Query Match 63.1%; Score 845.5; DB 1; Length 1758;
 Best Local Similarity 60.2%; Pred. No. 7.2e-77;
 Matches 147; Conservative 35; Mismatches 61; Indels 1; Gaps 1;
 QY 1 GLKRGSGSPAPWT-TRGFVTRHSQTAIPSCPESTVPLYSGFSLFYQGNRAHQ 59
 DB 1515 GLPGIPGPGSGMARPGRGTFPAKHSQTAIPSCPEASGLSLLYQGNRAHQ 1574
 QY 60 DLGLGSCGLRFTTPELFCVNDVNCFSRNDYSYWLSTPALMNNAPITGRALPEYI 119
 DB 1575 DLGPGSCLSKFTNTPFEMFCNNSVCHVSRNDYSYWLSTDEPMNPNVGTATRPYI 1634
 QY 120 SRCVCEBPALAIIVHSQTDIPPCPHWISLAKGFSINTFASGETGQALNSPGSCL 179
 DB 1635 SRCACEVPTIIIVHSQDTSVPQPGSGMWGTCYSFVHTTAGAEQTLQSLQ 1694
 QY 180 EEPFASPLECHGRGTGVYNSYSFNLASINPERMPKPIPSYVKADELKIIISRCQVC 239
 DB 1695 EEPFAPRIEDHGRTGVYNTNIGFWFSYVDQKQFRRPSQTLKAGLADRVSRQVC 1754
 QY 240 MKRR 243
 DB 1755 LKNR 1758

RESULT 8
 CA24_ASCSU STANDARD; PRT; 1763 AA.
 ID CA24_ASCSU
 AC P27393;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE COLLAGEN ALPHA 2(IV) CHAIN PRECURSOR.
 OS Ascaris suum (Pig roundworm) (Ascaris lumbricoides).
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
 OC Ascarididae; Ascaris.

```

OX NCBI_TaxID=6253;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91340768; PubMed-1714907;
RT Petlict J., Kingson I.B.;
RT "The complete primary structure of a nematode alpha 2(IV) collagen
RT and the partial structural organization of its gene.";
RL J. Biol. Chem. 266:16149-16156(1991).
CC -1- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.
CC -1- SUBUNIT: PRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.
CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH
CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NCI
CC DOMAINS.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; I (SHOWN HERE) AND II; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -1- P.TM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- P.TM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NCI DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M67507; AAA18014.1; -
DR PIR: S16366; S16366.
DR InterPro: IPR000087; -
DR InterPro: IPR001442; -
DR Pfam: PF01413; C4; 2.
DR Pfam: PF01391; Collagen; 23.
DR KM Hydroxylation: Connective tissue: Basement membrane; Repeat; Collagen;
DR Alternative splicing: Glycoprotein; Signal.
FT SIGNAL 1 26
FT CHAIN 27 1763
FT DOMAIN 27 42
FT DOMAIN 43 1529
FT DOMAIN 1530 1763
FT DISULFID 1548 1637
FT DISULFID 1561 1634
FT DISULFID 1593 1599
FT DISULFID 1656 1752
FT DISULFID 1690 1749
FT DISULFID 1702 1709
FT CARBOHYD 146 126
FT CARBOHYD 249 249
FT VARSPLIC 230 266
FT FT
FT SEQUENCE 1763 AA; 168526 MW; 304F528BC06AAED CRC64;
SQ
Query Match 59.4%; Score 796.5; DB 1; Length 1763;
Best Local Similarity 57.9%; Pred. No. 3, le-67;
Matches 140; Conservative 41; Mismatches 58; Indels 3; Gaps 2;
OY 1 GUKGRKDGSGPATWTRRGVFTFRHSQTALPSCREGTYPLVGFSEFLFYQGNORAHGDD 60
DB 1515 GLPGSGPGPGPGPGPKDGLLVKHSOTSEVPCPGPMKMGWYSLGLENGEKSHND 1574
OY 61 LSTLGSCLORFTTTPPLFCNVNDVNCFNASRNDYSYMLSTPALPMMAPIPTGALPEYIS 120
II ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

DB 1575 LGHAGSCLRSFSTMPPLFCDVNVNVCNYSRNDKSYWLSTTA--PIPMMPVSESGIEPYIS 1632
OY 121 RCTVCGSPALAIIVHSQTTDIPPCPHGWISLKGSEFMFTSAGSGTQOALASPSCLIE 180
DB 1633 RCACVCEAPAVAVIAVHSQTTIOPNCPGWNSLWIGYSFAMHTGAGAGGQOSLSSPSCLIE 1692
OY 181 EFASFLELCHG-RGTCNYSNSYSFWLMSINFERMRPPTSTVAGELEKTIKISQYOC 239
DB 1693 DPAATPFIECNARGKCHFRANKSFWLTITIEDQOFRIPESETLWAGSLRTVNSQYOC 1752
OY 240 MK 241
DB 1753 IR 1754
RESULT 9
CA24-CAEEL STANDARD; PRT; 1758 AA.
ID CA24-CAEEL
AC P17140;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE COLLAGEN ALPHA 2(IV) CHAIN PRECURSOR.
GN LET-2 OR CLB-1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RX MEDLINE-94012964; PubMed-7691828;
RA Sibley M.H., Johnson J.J., Mello C.C., Kramer J.M.;
RT "Genetic identification, sequence, and alternative splicing of the
RT Caenorhabditis elegans alpha 2(IV) collagen gene.";
RT J. Cell Biol. 123:255-264(1993).
RN [2]
RP PRELIMINARY SEQUENCE OF 1495-1758 FROM N.A.
RX STRAIN-BRISTOL N2;
RX MEDLINE-9008929; PubMed-2793871;
RA Guo X., Kramer J.M.;
RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
RT genes are located on separate chromosomes.";
RT J. Biol. Chem. 264:17574-17582(1989).
RN [3]
RP VARIANTS.
RX MEDLINE-94320591; PubMed-8045258;
RA Sibley M.H., Graham P.L., von Mende N., Kramer J.M.;
RT "Mutations in the alpha 2(IV) basement membrane collagen gene of
RT Caenorhabditis elegans produce phenotypes of differing severities.";
RT EMBO J. 13:3278-3285(1994).
CC -1- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.
CC -1- SUBUNIT: PRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.
CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH
CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NCI
CC DOMAINS.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; I (SHOWN HERE) AND II; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: FORM I IS PREDOMINANT IN EMBRYOS AND FORM
CC II IS PREDOMINANT IN THE LARVAE AND ADULTS.
CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -1- P.TM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- P.TM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NCI DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -1- DISEASE: MUTATIONS IN LET-2 ARE GENERALLY EMBRYONIC LETHAL.
CC -----

```


"Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer.";
 Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLomerular basement membranes (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEGLYCANS AND ENACTIN/
 CC NIDAGEN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: POLINES ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: M23334; AAA51626.1; -
 CC EMBL: M23333; AAA51626.1; JOINED.
 CC EMBL: J04695; AAA50293.1; -
 CC EMBL: J04648; AAA37438.1; -
 CC EMBL: X04647; CAA28308.1; -
 CC EMBL: M15833; AAA37341.1; -
 CC EMBL: X04410; CAA27998.1; -
 CC EMBL: X02896; CAA26655.1; -
 CC EMBL: X02897; CAA51614.1; -
 CC EMBL: X02898; CAA26657.1; -
 CC EMBL: X02899; CAA26658.1; -
 CC EMBL: M23334; AAA51627.1; -
 CC PIR: A33526; A33526.
 CC HSSP: P19712; 1KVD.
 CC MGD: MGI:88455; Col4a2.
 CC InterPro: IPR000087; -
 CC InterPro: IPR001442; -
 CC Pfam: PF01413; C4.2.
 CC Pfam: PF01391; Collagen; 19.
 CC KEGG: Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 CC Glycoprotein; Basement membrane; Collagen; Signal.
 CC KW SIGNAL.
 CC FT PROPEP 1 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
 CC FT CHAIN 184 1707 COLLAGEN ALPHA 2(IV) CHAIN.
 CC FT DOMAIN 184 1479 TRIPLE-HELICAL REGION.
 CC FT DISULFID 1499 1588 NONHELICAL REGION (NC1).
 CC FT DISULFID 1532 1585 OR 1585 (BY SIMILARITY).
 CC FT DISULFID 1544 1550 OR 1588 (BY SIMILARITY).
 CC FT DISULFID 1607 1703 BY SIMILARITY.
 CC FT DISULFID 1641 1700 OR 1703 (BY SIMILARITY).
 CC FT DISULFID 1653 1660 BY SIMILARITY.
 CC FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 1270 1270 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CONFLICT 1051 1051 P -> R (IN REF. 6).
 CC FT CONFLICT 1097 1097 S -> G (IN REF. 7).
 CC FT CONFLICT 1171 1171 G -> S (IN REF. 6).
 CC FT CONFLICT 1179 1179 P -> R (IN REF. 6).
 CC FT CONFLICT 1241 1241 Q -> E (IN REF. 6).
 CC FT CONFLICT 1328 1328 P -> A (IN REF. 6).
 CC FT CONFLICT 1573 1573 V -> L (IN REF. 4).
 CC FT CONFLICT 1623 1623 Y -> H (IN REF. 4).
 CC FT SEQUENCE 1707 AA: 167391 MW: 1A565159605FD508 CRC64;

Query Match 56.8%; Score 760.5; DB 1; Length 1707;
 Best Local Similarity 57.6%; Pred. No. 7.5e-64;
 Matches 140; Conservative 37; Mismatches 61; Indels 5; Gaps 4;
 QY 1 GLKRGDGSQSPATWTRGEPFTRSHQSTTAIDSPCEGYPLISGFSEFLVQGNRAHGD 60
 DB 1466 GRGSGGLGMPGRSISGLTLVHKSHQTOEPGCPGMKMLWSGLYLFEGGEKHAND 1525
 QY 61 LGTLGSCLOREFTMPPLFCNVDCNCFASRNDYSYVLSPTALPMMAITGRALPEYIS 120
 DB 1526 LGLASGLAFSTMPPLFCNVDCNCFASRNDYSYVLSPTALPMMAITGRALPEYIS 1583
 QY 121 RCTVCEGPAIALAVHSQTDIPPCPHGWISLWKGFSIFMTSGSEGTQALASPGSCLE 180
 DB 1584 RGVCEAPVALVAVHSQTDIPPCPHGWISLWKGFSIFMTSGSEGTQALASPGSCLE 1643
 QY 181 EFRASPFLECH-GRGICNTYSNSTFWLASLNPERRFR-PIPSYKAGELEKIRCOV 238
 DB 1644 DFRATPEFIEGNGRGTCYHFAFKYFWLFTTI-PEONFOSPTSPADTLKAGLIRTHISRCV 1702
 QY 239 CMK 241
 DB 1703 CMK 1705
 RESULT 11
 CA24.HUMAN STANDARD; PRT; 1712 AA.
 ID CA24.HUMAN
 AC P08572;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE COLLAGEN ALPHA 2(IV) CHAIN PRECURSOR.
 GN COL4A2
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP MEDLINE=89066769; PubMed=3198637;
 RA Hostlika S.L., Tryggvason K.;
 RT "The complete primary structure of the alpha 2 chain of human type IV
 RL J. Biol. Chem. 263:19488-19493(1988).
 RN [2]
 RP SEQUENCE OF 1-1042 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=88151998; PubMed=3345760;
 RA Brazel D., Pollner R., Oberhauser I., Kuehn K.;
 RT "Human basement membrane collagen (type IV). The amino acid sequence
 RL Eur. J. Biochem. 172:35-42(1988).
 RN [3]
 RP SEQUENCE OF 1254-1712 FROM N.A.
 RX MEDLINE=87219158; PubMed=3582677;
 RA Hostlika S.L., Kurkunen M., Tryggvason K.;
 RT "Nucleotide sequence coding for the human type IV collagen alpha 2
 RL chain cDNA reveals extensive homology with the NC-1 domain of alpha 1
 FEBS Lett. 216:281-286(1987).
 RN [4]
 RP SEQUENCE OF 1451-1485 FROM N.A.
 RX MEDLINE=87092458; PubMed=3025878;
 RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;
 RT "Human collagen genes encoding basement membrane alpha 1 (IV) and
 RL alpha 2 (IV) chains map to the distal long arm of chromosome 13.";
 Proc. Natl. Acad. Sci. U.S.A. 84:512-516(1987).
 RN [5]
 RP SEQUENCE OF 1486-1712 FROM N.A.

RX MEDLINE-87250571; PubMed-2439508;
 RA Myers J.C., Howard P.S., Jelen A.M., Macarak E.J.;
 RT "Duplication of type IV collagen COOH-terminal repeats and species-
 RT specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";
 RL J. Biol. Chem. 262:9231-9238(1987).
 RN [6]
 RP SEQUENCE OF 1-33 FROM N.A.
 RX MEDLINE-89034231; PubMed-3182844;
 RA Sothman R., Houtari M., Hostilka S.L., Prockop D.J., Trygvason K.;
 RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
 RT collagen are divergently encoded on opposite DNA strands and have an
 RT overlapping promoter region.";
 RL J. Biol. Chem. 263:17217-17220(1988).
 RN [17]
 RP SEQUENCE OF 1-33 FROM N.A.
 RX MEDLINE-89030632; PubMed-2846280;
 RA Poeschl E., Pollner R., Kuehn K.;
 RT "The genes for the alpha 1(IV) and alpha 2(IV) chains of human
 RT basement membrane collagen type IV are arranged head-to-head and
 RT separated by a bidirectional promoter of unique structure.";
 RL EMBO J. 7:2687-2695(1988).
 RN [8]
 RP SEQUENCE OF 1-33 FROM N.A.
 RX TISSUE-Skin;
 RC MEDLINE-93305049; PubMed-8317999;
 RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;
 RT "Identification of a novel sequence element in the common promoter
 RT region of human collagen type IV genes, involved in the regulation of
 RT divergent transcription.";
 RL Biochem. J. 292:687-695(1993).
 RN [9]
 RP SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.
 RX TISSUE-Placenta;
 RC MEDLINE-89005112; PubMed-2844531;
 RA Stehob B., Deutzmann R., Kuehn K.;
 RT "The arrangement of intra- and intermolecular disulfide bonds in the
 RT carboxyterminal, non-collagenous aggregation and cross-linking domain
 RT of basement-membrane type IV collagen.";
 RL Eur. J. Biochem. 176:617-624(1988).
 RL [1]
 RP FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 CC NIDOGEN.
 CC [1]
 RP SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOPFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC [1]
 RP DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL, TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC [1]
 RP PPM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC [1]
 RP PPM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC [1]
 RP This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC [1]
 RP EMBL: X05562; CAA29076.1; -
 DR EMBL: X05610; CAA29098.1; -
 DR EMBL: J02760; AAA58422.1; -
 DR EMBL: M36963; AAA53099.1; -
 DR EMBL: X12784; CAA31275.1; -
 DR EMBL: J04217; AAA53097.1; -
 DR PIR: A32024; A32024.

DR MM; 120090;
 DR InterPro: IPR000087;
 DR InterPro: IPR001442;
 DR Pfam: PF01413; C4; 2.
 DR Pfam: PF01391; Collagen; 20.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Basement membrane; Collagen; Signal.
 FT SIGNAL 1 25
 FT PROPEP 26 183 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).
 FT CHAIN 184 1712 COLLAGEN ALPHA 2(IV) CHAIN.
 FT DOMAIN 184 1484 TRIPLE-HELICAL REGION.
 FT DOMAIN 1485 1712 NONHELICAL REGION (NC1).
 FT DISULFID 1504 1593 OR 1590 (BY SIMILARITY).
 FT DISULFID 1537 1590 OR 1593 (BY SIMILARITY).
 FT DISULFID 1549 1555 BY SIMILARITY.
 FT DISULFID 1612 1708 OR 1705 (BY SIMILARITY).
 FT DISULFID 1646 1705 OR 1708 (BY SIMILARITY).
 FT DISULFID 1658 1665 BY SIMILARITY.
 FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .).
 FT CONFLICT 471 471 R -> P (IN REF. 2).
 FT CONFLICT 683 683 A -> G (IN REF. 2).
 FT CONFLICT 1575 1575 M -> I (IN REF. 5).
 FT CONFLICT 1663 1663 G -> H (IN REF. 9).
 FT CONFLICT 1701 1701 H -> G (IN REF. 9).
 SQ SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;
 Query Match 56.4%; Score 755.5; DB 1; Length 1712;
 Best Local Similarity 58.0%; Pred. No. 2,2e-63;
 Matches 141; Conservative 34; Mismatches 63; Indels 5; Gaps 4;
 Db 1 GLKRGSGSPATWTFGFTVTRHSQTATPSCPEGVPLXSGSFVFGNGRAHOD 60
 1471 GRPSPGLPGMRGSRVSIGYLLVHRSOTDQEPMPGVGNKMLMSGSLYFPGQEXAHND 1530
 61 LGTASGLQRTTPTFELCNVDVCFNFSRNDYSWTSLPALMNMAMPITGRALEPYIS 120
 1531 LGLAGSCLARSTPELXCNQDVCYVARSNDKSWLSTTA--PLPMNPAVDEDEKPYIS 1588
 121 RCYVCEGPAIVAVHSQTTDIPPCPHGWSLAKGFSFLMTFSAGSEGTGQALASPGSLE 180
 1589 RSCVCEAPAIIVAVHSQDVSIPHCPCGMRSLIGYSLMTHRAAGDEGCGOSLVSPGSCLE 1648
 181 EFRASPFECH-GRGTCVYNSYSFWLASLNPSEMER-KPLPSVYKGELEKITSROY 238
 1649 DFRATPTECGMGKCHYVANKYSFWLTTI-PRDSFGSPSADTLKAGLIRTHISROY 1707
 Db 239 CMK 241
 1708 CMK 1710
 RESULT 12
 CA64 HUMAN STANDARD; PRT; 1678 AA.
 ID CA64 HUMAN 014031;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE COLLAGEN ALPHA 6(IV) CHAIN PRECURSOR.
 GN COL4A6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Eye, and Kidney;
 RX MEDLINE-94171779; PubMed-8125972;
 RA Ohashi T., Sugimoto M., Mattei M.-G., Ninomiya Y.;
 RT "Identification of a new collagen IV chain, alpha 6(IV), by cDNA
 RT isolation and assignment of the gene to chromosome Xq22, which is the
 RT same locus for COL4A5.";

RL J. Biol. Chem. 269:7520-7526(1994).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 CC NIDOGEN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- SUPRACELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D21337; BAA04809.1; .
 DR MIM: 303631; .
 DR InterPro: IPR000087; .
 DR InterPro: IPR001442; .
 DR Pfam: PF01413; C4.2.
 DR Pfam: PF01391; Collagen; 21.
 KM Extracellular matrix; Connective tissue; Basement membrane;
 Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Collagen; Signal.
 FT SIGNAL 1 15
 FT CHAIN 16 1678
 FT DOMAIN 16 45
 FT DOMAIN 46 1450
 FT DOMAIN 1451 1678
 FT SITE 514 516
 FT SITE 559 561
 FT SITE 985 987
 FT DISULFID 1469 1558
 FT DISULFID 1502 1553
 FT DISULFID 1514 1520
 FT DISULFID 1577 1674
 FT DISULFID 1611 1671
 FT DISULFID 1623 1630
 FT CARBOHYD 126 126
 FT SEQUENCE 1678 AA; 162461 MW; 6564E282D3D7BD CRC64.
 SO
 Query Match 55.1%; Score 739; DB 1; Length 1678;
 Best Local Similarity 54.7%; Pred. No. 7.9e-62;
 Matches 133; Conservative 42; Mismatches 64; Indels 4; Gaps 3;

Db 1614 DFRATPEICSGARGTCHYFANKYSFWLTVVERQKGELPVSSETLAKAGLHTRVRSQV 1673
 QY 239 CMK 241
 Db 1674 CMK 1676
 RESULT 13
 CA44-BOVIN STANDARD; PRT: 453 AA.
 ID CA44-BOVIN
 AC 029442; 1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE COLLAGEN ALPHA 4(IV) CHAIN (FRAGMENT).
 CN COL4A4
 OS Bos taurus (bovine); Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Eukaryota; Metazoa; Chordata; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 317-328.
 RC TISSUE=Lens;
 RA MEDLINE=92112769; PubMed=1370461;
 RA Mariyama M., Kalluri R., Hudson B.G., Reeders S.T.;
 RT "The alpha 4(IV) chain of basement membrane collagen. Isolation of
 RT cDNAs encoding bovine alpha 4(IV) and comparison with other type IV
 RT collagens.";
 RT J. Biol. Chem. 267:1253-1258(1992).
 RN [2]
 RP SEQUENCE OF 217-246.
 RX MEDLINE=90202779; PubMed=2318822;
 RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
 RT "Glomerular basement membrane. Identification of a fourth chain,
 RT alpha 4, of type IV collagen.";
 RT J. Biol. Chem. 265:5466-5469(1990).
 RN [3]
 RP SEQUENCE OF 217-233.
 RX MEDLINE=87222419; PubMed=2438283;
 RA Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
 RA Hudson B.G.;
 RT "Localization of the Goodpasture epitope to a novel chain of basement
 RT membrane collagen.";
 RL J. Biol. Chem. 262:7874-7877(1987).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 CC NIDOGEN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- SUPRACELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
 CC COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
 CC COCHLEA, LUNG AND BRAIN.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: M77480; AAA30458.2; ALT_SEQ.

DR InterPro: IPR000087; -

DR InterPro: IPR001442; -

DR Pfam: PF01413; C4; 2.

DR Pfam: PF01391; Collagen; 3.

KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation.

KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.

FT NON_TER 1

FT DOMAIN <1 222

FT DISULFID 223 453

FT DISULFID 243 332

FT DISULFID 276 329

FT DISULFID 288 294

FT DISULFID 351 449

FT DISULFID 385 446

FT DISULFID 397 404

FT CONFLICT 219 219

FT SEQUENCE 453 AA; 46384 MW; F7ED410AE9A65BC1 CRC64;

Query Match Best Local Similarity 52.9%; Score 709; DB 1; Length 453;
Matches 130; Conservative 42; Mismatches 67; Indels 22; Gaps 5;

QY 1 GLKGRDSDGSPAT-----WTRGFVFTNHSOTTAIPSCPEGTPLYS 43

DB 193 GHKDMGEAGCGAPGPPGPGDPGPGIGFGYLSGLFLVLSQTDGPPTCPMGPRMT 252

QY 44 GSFLEFVQGNRAHGODLTGSCLORTTMTPLFCNVNDVNCASRDYTWLSTPALM 103

DB 253 GSYLTYLEGGERAHNODGLAGSCLPTFTLPFAVCNHYCHYCARNDKSTWLASAG-- 310

QY 104 PMNAPITGRALPEYISRCTVCEGPATIAIAVHSOTTDIPPCPHGMSISLKGFSFIMPTSA 163

DB 311 PLPMPLSEEDIRYISCAVCEAPQAVVHSODQSIIPCPRAMRSIMWISFLMHTGA 370

QY 164 GSEGTGALASPGSCLFEFRASPLECHGR-GTCNYNSNSYFWLASINPEFRK-PIP 221

DB 371 GDGGGQALMSPGSCLDEFRAPLECGRGCTGFHFAFKKSFMLTVPDLQVFSAPSP 430

QY 222 STVAGELEK-IISRCOVCKM 241

DB 431 DTLKESHAOROKISRCOVCK 451

RESULT 14

CA44_RABIT STANDARD; PRT; 623 AA.

AC P53787; 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE COLLAGEN ALPHA 4(IV) CHAIN (FRAGMENT).

GN COL4A4.

OS Oryctolagus cuniculus (Rabbit).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI_TaxID=9986;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Corneal endothelium;

RX MEDLINE=93054733; PubMed=1429714;

RA Kamagata Y., Mattei M.-G., Ninomiya Y.;

RT Isolation and sequencing of cDNAs and genomic DNAs encoding the

RT alpha 4 chain of basement membrane collagen type IV and assignment of

RT the gene to the distal long arm of human chromosome 2.;

RT J. Biol. Chem. 267:23753-23758(1992).

CC -!- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF

CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'

CC MESHWORK TOGETHER WITH LAMININS, PROTEGLYCANS AND ENACTIN/
CC NIDOGN.

CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE

CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.

CC -!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).

CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS

CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE

CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY

CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL

CC TRIPLE-HELICAL 7S DOMAIN.

CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING

CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

CC -!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH

CC ARE INVOLVED IN INTRA- AND INTERMOLECULAR DISULFIDE BONDING. 12 OF

CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE

CC IV COLLAGENS.

CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: L01477; -; NOT_ANNOTATED_CDS.

DR InterPro: IPR000087; -

DR InterPro: IPR001442; -

DR Pfam: PF01413; C4; 2.

DR Pfam: PF01391; Collagen; 5.

KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;

KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.

FT NON_TER 1

FT DOMAIN <1 392

FT DISULFID 393 623

FT DISULFID 413 502

FT DISULFID 446 499

FT DISULFID 458 464

FT DISULFID 521 619

FT DISULFID 555 616

FT DISULFID 567 574

FT SEQUENCE 623 AA; 62393 MW; CCB93BB31242FBE2 CRC64;

Query Match Best Local Similarity 52.4%; Score 702; DB 1; Length 623;
Matches 128; Conservative 45; Mismatches 66; Indels 22; Gaps 5;

QY 1 GLKGRDSDGSPAT-----WTRGFVFTNHSOTTAIPSCPEGTPLYS 43

DB 363 GHKDDTGEAGCGAPGPPGPGDPGPGKIGFGYLSGLFLVLSQTDGPACPMGPRMT 422

QY 44 GSFLEFVQGNRAHGODLTGSCLORTTMTPLFCNVNDVNCASRDYTWLSTPALM 103

DB 423 GSYLTYLEGGERAHNODGLAGSCLPTFTLPFAVCNHYCHYCARNDKSTWLASAG-- 480

QY 104 PMNAPITGRALPEYISRCTVCEGPATIAIAVHSOTTDIPPCPHGMSISLKGFSFIMPTSA 163

DB 481 PLPMPLSEEDIRYISCAVCEAPQAVVHSODQSIIPCPRAMRSIMWISFLMHTGA 540

QY 164 GSEGTGALASPGSCLFEFRASPLECHGR-GTCNYNSNSYFWLASINPEFRK-PIP 221

DB 541 GDGGGQALMSPGSCLDEFRAPLECGRGCTGFHFAFKKSFMLTVPDLQVFSAPSP 600

QY 222 STVAGELEK-IISRCOVCKM 241

DB 601 DTLKESHAOROKISRCOVCK 621

RESULT 15

CA44_HUMAN STANDARD: PRT; 1690 AA.
 ID CA44_HUMAN
 AC P53420:
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE COLLAGEN ALPHA 4(IV) CHAIN PRECURSOR.
 GN COL4A4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=9501445; PubMed=7523402;
 RA Leinonen A., Mariyama M., Mochizuki T., Tryggvason K., Reeders S.T.;
 RT "Complete primary structure of the human type IV collagen alpha 4(IV)
 chain. Comparison with structure and expression of the other alpha
 (IV) chains.";
 RJ J. Biol. Chem. 269:26172-26177(1994).
 [2]
 RP SEQUENCE OF 1219-1690 FROM N.A.
 RC TISSUE=Eye;
 RX MEDLINE=9337407; PubMed=8365481;
 RA Sugimoto M., Ohashi T., Yoshioka H., Matsuo N., Niinomiya Y.;
 RT "cDNA isolation and partial gene structure of the human alpha 4(IV)
 collagen chain.";
 RJ FEBS Lett. 330:122-128(1993).
 [3]
 RP SEQUENCE OF 1407-1507 FROM N.A.
 RX MEDLINE=9305473; PubMed=1429714;
 RA Kamagata Y., Mattei M.-G., Niinomiya Y.;
 RT "Isolation and sequencing of cDNAs and genomic DNAs encoding the
 alpha 4 chain of basement membrane collagen type IV and assignment of
 the gene to the distal long arm of human chromosome 2.";
 RJ J. Biol. Chem. 267:23753-23758(1992).
 [4]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97338662; PubMed=9195222;
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
 RT "The clinical spectrum of type IV collagen mutations.";
 RJ Hum. Mutat. 9:477-499(1997).
 [5]
 RP VARIANT AS SER-1201.
 RX MEDLINE=95078927; PubMed=7987396;
 RA Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Gubler M.-C.,
 RA Smeets H.J., Reeders S.T.;
 RT "Identification of mutations in the alpha 3(IV) and alpha 4(IV)
 collagen genes in autosomal recessive Alport syndrome.";
 RJ Nat. Genet. 8:77-82(1994).
 [6]
 RP VARIANT BPH GIU-897.
 RX MEDLINE=96379660; PubMed=8787673;
 RA Lemmink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,
 RA Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;
 RT "Benign familial hematuria due to mutation of the type IV collagen
 alpha4 gene.";
 RJ J. Clin. Invest. 98:1114-1118(1996).
 [7]
 RP VARIANTS AS, AND VARIANTS.
 RX MEDLINE=99011253; PubMed=9792860;
 RA Boye E., Mollet G., Forestier L., Cohen-Solal L., Helidet L.,
 RA Cochot P., Gruenfeld J.-P., Palcoux J.-B., Gubler M.-C., Antignac C.;
 RT "Determination of the genomic structure of the COL4A4 gene and of
 novel mutations causing autosomal recessive Alport syndrome.";
 RJ Am. J. Hum. Genet. 63:1329-1340(1998).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 CC NIDOGEN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-

CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
 CC COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
 CC COCHLEA, LUNG AND BRAIN.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPLEPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -1- DISEASE: DEFECTS IN COL4A4 ARE ASSOCIATED WITH THE TYPE II
 CC AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME (AS), AN HEREDITARY
 CC GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,
 CC HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN
 CC MALES AND FEMALES.
 CC -1- DISEASE: DEFECTS IN COL4A4 ARE ASSOCIATED WITH FAMILIAL BENIGN
 CC HEMATURIA (FBH) OR THIN BASEMENT MEMBRANE DISEASE. FBH IS
 CC CHARACTERIZED BY PERSISTENT HEMATURIA, AN ELECTRON MICROSCOPICALLY
 CC DETECTABLE THIN GLOMERULAR BASEMENT MEMBRANE (GBM) AND AN
 CC AUTOSOMAL DOMINANT MODE OF INHERITANCE. RENAL FUNCTION REMAINS
 CC NORMAL. IN CHILDREN DIFFERENTIATION BETWEEN FBH AND AS CAN BE
 CC DIFFICULT, BECAUSE BOTH DISORDERS ARE MANIFESTED BY PERSISTENT
 CC HEMATURIA AND THIN GBM AT THAT AGE.
 CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X81053; CAA56943.1; -
 CC EMBL: D17391; BAA04214.1; -
 CC DR MIM: 120131; -
 CC DR MIM: 141200; -
 CC DR MIM: 203780; -
 CC DR InterPro: IPR000087; -
 CC DR InterPro: IPR001442; -
 CC DR Pfam: PF01413; C4.2; -
 CC DR Pfam: PF01391; Collagen; 20.
 CC KW Extracellular matrix; Connective tissue; Basement membrane; Repeat;
 CC KW Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;
 CC KW Polymorphism.
 CC FT SIGNAL 1 38
 CC FT CHAIN 39 1690
 CC FT DOMAIN 39 64
 CC FT DOMAIN 65 1459
 CC FT DOMAIN 1460 1690
 CC FT SITE 94 96
 CC FT SITE 145 147
 CC FT SITE 169 191
 CC FT SITE 310 312
 CC FT SITE 724 726
 CC FT SITE 785 787
 CC FT SITE 989 991
 CC FT SITE 1206 1206
 CC FT SITE 1212 1214
 CC FT SITE 1480 1569
 CC FT DISULFID 1513 1566
 CC FT DISULFID 1525 1531
 CC FT DISULFID 1588 1686
 CC FT DISULFID 1622 1683
 CC FT DISULFID 1634 1641
 CC
 CC POTENTIAL.
 CC COLLAGEN ALPHA 4(IV) CHAIN.
 CC 7S DOMAIN.
 CC TRIPLE-HELICAL REGION.
 CC NONHELICAL REGION (NC1).
 CC CELL ATTACHMENT SITE (POTENTIAL).
 CC CELL ATTACHMENT SITE (POTENTIAL).
 CC CELL ATTACHMENT SITE (POTENTIAL).
 CC CELL ATTACHMENT SITE (POTENTIAL).
 CC CELL ATTACHMENT SITE (POTENTIAL).
 CC CELL ATTACHMENT SITE (POTENTIAL).
 CC CELL ATTACHMENT SITE (POTENTIAL).
 CC CELL ATTACHMENT SITE (POTENTIAL).
 CC CLEAVAGE (BY COLLAGENASE)
 CC (BY SIMILARITY).
 CC CELL ATTACHMENT SITE (POTENTIAL).
 CC OR 1566 (BY SIMILARITY).
 CC OR 1569 (BY SIMILARITY).
 CC BY SIMILARITY.
 CC OR 1683 (BY SIMILARITY).
 CC OR 1686 (BY SIMILARITY).
 CC BY SIMILARITY.


```
FT CARBOHYD 142 142 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 441 446 MISSING (IN AS).
FT VARIANT 545 545 /FTID-VAR_008148.
FT VARIANT 545 545 G -> A.
FT VARIANT 570 570 /FTID-VAR_008149.
FT VARIANT 570 570 E -> O.
FT VARIANT 897 897 /FTID-VAR_008150.
FT VARIANT 897 897 G -> E (IN PBH).
FT VARIANT 931 931 /FTID-VAR_001912.
FT VARIANT 931 931 A -> T.
FT VARIANT 1004 1004 /FTID-VAR_008151.
FT VARIANT 1004 1004 L -> P.
FT VARIANT 1030 1030 /FTID-VAR_008152.
FT VARIANT 1030 1030 G -> V (IN AS).
FT VARIANT 1201 1201 /FTID-VAR_008153.
FT VARIANT 1201 1201 G -> S (IN AS).
FT VARIANT 1402 1402 /FTID-VAR_001913.
FT VARIANT 1402 1402 P -> S.
FT VARIANT 1572 1572 /FTID-VAR_008154.
FT VARIANT 1572 1572 P -> L (IN AS).
FT CONFLICT 1659 1660 /FTID-VAR_008155.
FT CONFLICT 1659 1660 LQ -> FE (IN REF. 2).
SQ SEQUENCE 1690 AA; 164095 MM; E1E72F283A72BAAE CRC64;
```

Query Match 51.9%; Score 696; DB 1; Length 1690;
Best Local Similarity 49.0%; Pred. No. 9.2e-58;
Matches 128; Conservative 43; Mismatches 68; Indels 22; Gaps 5;

```
QY 1 GLGKRGDSGSP-----ATWTRGFVETRHSTQTAIPSCSEGTVPPLYS 43
DB 1430 GRKGDYGEDEYPCGPGRPGPIGDPGRGFGPGYLGLVLVHSTQDQPTCLGMPRLWT 1489
QY 44 GSFLEFVQGNORAHGODLGTLSCLQRTTMTPELCNVNDVCNFPASRNDYSYWLSTPALM 103
DB 1490 GYSLLYLEGEKAKHNDLGLAGSCLPVEFTLPAYCNHQCVCYQARNDRSYWLASAA-- 1547
QY 104 PMMAPITGALPPIYISRCTVCEGPATIAIVHSQTDIPPCPHGMISLMKGFSLMFTSA 163
DB 1548 PLPMMPLESEAIRPYVSRCAVCEAPAAVAVHSDOSIPPCPQTRSLMIGYSFLMHTGA 1607
QY 164 GSEGTQALASPGSCLEEFRAFPLECHGR-GTCNYSNSYSFWLASLNPBRMFRK-PIP 221
DB 1608 GDGGGGQALMSPGSCLEDFRAAFPLECQGRGCTCHFRANKYSFWLTVKADLPSSAPAR 1667
QY 222 STYKAGLEK-TISRCQVCMK 241
DB 1668 DTLKESQAROKISRQCVCMK 1688
```

Search completed: April 13, 2001, 15:46:04
Job time: 277 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2001, 15:38:59 ; Search time 33.96 Seconds

(Without alignments)
493.769 Million cell updates/sec

Title: US-09-543-371-10

Perfect score: 1340
Sequence: 1 GLKGRGSGSPATWTTRGF.....KAGELEKISRQVCMKKRH 244

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 0%
Listing first 45 summaries

Database: PIR-67:*
1: pirl:*
2: pirl:*
3: pirl:*
4: pirl:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1340	100.0	1670	1 CGH03B	collagen alpha 3(I
2	1218.5	90.9	246	2 I48302	collagen alpha 3(I
3	1210.5	90.3	471	2 A39024	collagen alpha 3(I
4	960	71.6	1659	1 CGH04B	collagen alpha 1(I
5	952.5	71.1	253	2 I48304	collagen alpha 5(I
6	951	71.0	1659	1 CGM54B	collagen alpha 1(I
7	943.5	70.4	1691	1 S22917	collagen alpha 1(I
8	943	70.4	258	2 B61228	collagen alpha 1(I
9	922.5	65.1	754	2 A55267	collagen alpha 1(I
10	872	65.1	220	2 B49736	collagen alpha 3(I
11	852.5	63.6	1744	2 S40991	collagen alpha 1(I
12	834	62.2	161	2 S49488	collagen alpha 3(I
13	832	62.1	1752	2 A45407	collagen alpha 3(I
14	796.5	59.4	1753	2 S16366	collagen alpha 2(I
15	783.5	58.5	1758	2 T29350	hypothetical prote
16	783.5	58.5	1759	2 T29351	collagen alpha 2(I
17	782	58.4	261	2 A34476	collagen alpha 2(I
18	767.5	57.3	1747	2 A54121	collagen alpha-4 c
19	760.5	56.8	1707	2 A33526	collagen alpha-4 c
20	755.5	56.4	1712	1 CGH02B	collagen alpha 2(I
21	739	55.1	1691	1 CGH06B	collagen alpha 2(I
22	734.5	54.8	775	2 A61228	collagen alpha 6(I
23	709	52.9	453	2 S18804	collagen alpha 2(I
24	705	52.6	312	2 I48303	collagen alpha 4(I
25	702	52.4	623	2 A45137	collagen alpha 4(I
26	696	51.9	1690	1 CGH01B	collagen alpha 4(I
27	694.5	51.8	1775	2 A31893	collagen alpha 1(I
28	646.5	48.2	1761	2 T13990	collagen type IV a
29	334	24.9	81	2 A49736	collagen alpha 3(I

30	326	24.3	79	2 C43928	probable collagen
31	228	17.0	58	2 B43928	probable collagen
32	198	14.8	52	2 S69113	collagen alpha 3(I
33	110.5	8.2	260	2 T23033	hypothetical prote
34	96.5	7.2	4391	2 A38096	hypothetical prote
35	90.5	6.8	2150	2 T32497	perlecan precursor
36	89.5	6.7	537	2 S78195	hypothetical prote
37	89.5	6.7	573	2 S78197	cytochrome-c oxida
38	89.5	6.7	3707	2 S18252	probable maturase
39	84.5	6.3	369	2 G72069	heparan sulfate pr
40	84	6.3	269	2 D69223	conserved hypotnet
41	82.5	6.2	1016	2 T00375	anion transport sy
42	81.5	6.1	539	2 A35052	hypothetical prote
43	81.5	6.1	2531	2 T16743	interleukin-2 rece
44	81	6.0	423	2 A41207	hypothetical prote
45	81	6.0	599	2 B42049	collagen 13, nonfi
					telstmanolysin (EC

ALIGNMENTS

RESULT 1
CGH03B
collagen alpha 3(IV) chain precursor, long splice form - human
N:Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice for
C:Species: Homo sapiens (man)
C:Date: 28-Oct-1994 #sequence-revision 03-Oct-1995 #text-change 22-Jun-1999
C:Accession: A54763, A43928, A44043, A45921, A30786
J:Maruyama, M.; Leinonen, A.; Mochizuki, T.; Trygvaeson, K.; Reeder, S.T.
J. Biol. Chem. 269, 23013-23017, 1994
A:Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpress
A:Reference number: A54763; MUID:94364994
A:Accession: A54763
A:Molecule type: mRNA
A:Residues: 1-1670 <MAR>
A:Cross-references: GB:X80031; NID:9577563; PID:9577564
A:Experimental source: kidney
R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
J. Clin. Invest. 89, 592-601, 1992
A:Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be th
A:Reference number: A43928; MUID:92147878
A:Accession: A43928
A:Molecule type: mRNA
A:Residues: 1311-1524, 1, 1526-1670 <TUR>
A:Cross-references: GB:M81379
A:Experimental source: kidney
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 267, 19780-19784, 1992
A:Title: Exon/intron structure of the human alpha 3(IV) gene encompassing the Goodpa
ction.
A:Reference number: A44043; MUID:93015826
A:Accession: A44043
A:Molecule type: DNA; mRNA
A:Residues: 1386-1670 <QUI>
A:Cross-references: GB:M92993; NID:9177895; PID:NAA21610.1; PID:9177896
A:Note: sequence extracted from NCBI backbone (NCBI:115597)
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 269, 17358, 1994
A:Reference number: A44738; MUID:94274734
R:Contents: annotation; erratum; correction to intronic sequence in A44043
R:Bernal, D.; Quinones, S.; Saus, J.
J. Biol. Chem. 268, 12090-12094, 1993
A:Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.
A:Reference number: A45971; MUID:93280184
A:Accession: A45971
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1427-1444 <BER>
A:Note: sequence extracted from NCBI backbone (NCBI:13363); sequence incorrectly id
Am. J. Hum. Genet. 49, 545-554, 1991
A:Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain
A:Reference number: A39786; MUID:91353570

A:Accession: A39786
 A:Molecule type: mRNA
 A:Residues: 1453-1593, 'A', 1595-1670 <MOR>
 A:Cross-references: GB:S55790; NID:9234418; PIDN:AA19637.1; PID:9234419
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 ed and subsequently O-glycosylated.
 C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope
 C:Genetics:
 A:Gene: GDB:COL4A3
 A:Cross-references: GDB:128351; OMIM:120070
 A:Map position: 2q36-2q37
 A:Introns: 1385/1; 1418/1; 1488/1; 1547/2; 1585/3; 1643/2 #status incomplete
 A:Notes: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
 C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3
 monomer amino-terminal domains (with disulfide and desmosine cross-links), dimeric 3
 er associations in the interrupted helical domain (with disulfide and desmosine cross-
 C:Function:
 A:Description: minor structural component of extracellular basement membrane in kidney
 C:Superfamily: collagen alpha 1(IV) chain
 Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel
 l; 28/Domain: signal sequence #status predicted <Sig>
 F:3-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <Mat>
 F:29-42/Domain: amino-terminal nonhelical, NH1 <NH1>
 F:43-1438/Region: interrupted helical
 F:791-793/Region: cell attachment (R-G-D) motif
 F:996-998/Region: cell attachment (R-G-D) motif
 F:1154-1156/Region: cell attachment (R-G-D) motif
 F:1306-1308/Region: cell attachment (R-G-D) motif
 F:1345-1347/Region: cell attachment (R-G-D) motif
 F:1432-1434/Region: cell attachment (R-G-D) motif
 F:1439-1670/Domain: collagen IV carboxyl-terminal nonhelical, NC1 <NC1>
 F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CR1>
 F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CR2>
 F:31,33,41,125,422,476,479,682,722,809,1387/Disulfide bonds: interchain #status predi
 F:253/Binding site: carboxydirate (Asn) (covalent) #status predicted
 F:1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
 F:1505-1511,1616-1622/Disulfide bonds: #status predicted
 F:1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 100.0%; Score 1340; DB 1; Length 1670;
 Best Local Similarity 100.0%; Pred. No. 1.5e-113;
 Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GLKRGDSSPATWTRGFEVTRHSQTTATPSCPEGVPLYSGFSEFLVQGNRAHQD 60
 1427 GLKRGDSSPATWTRGFEVTRHSQTTATPSCPEGVPLYSGFSEFLVQGNRAHQD 1486
 61 LGTIGSCLQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALMNMNAPITGRALPEYIS 120
 1487 LGTIGSCLQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALMNMNAPITGRALPEYIS 1546
 QY 121 RCYVCEGPAIAIVHSQTTDIPCPCHGWSLWKGFSTLMTFSASSEGQALASPGSCL 180
 1547 RCYVCEGPAIAIVHSQTTDIPCPCHGWSLWKGFSTLMTFSASSEGQALASPGSCL 1606
 QY 181 EFRASPLECHRGRCNCYNSYSFWLNLNPERFRKPISTYKAGLEKIIISRCQYCM 240
 1607 EFRASPLECHRGRCNCYNSYSFWLNLNPERFRKPISTYKAGLEKIIISRCQYCM 1666
 QY 241 KRRH 244
 1667 KRRH 1670

RESULT 2
 collagen alpha 3(IV) chain - mouse (fragment)
 C:Species: Mus musculus (house mouse)
 C:Date: 15-Mar-1996 #sequence_revision 15-Mar-1996 #text_change 16-Feb-1997
 C:Accession: 148302; S47278
 R:Miner, J.H.; Sanes, J.R.
 J. Cell Biol. 127, 879-891, 1994

A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae; se
 A:Reference number: A54979; MUID:95050957
 A:Accession: 148302
 A>Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-246 <RES>
 A:Cross-references: EMBL:Z35166; NID:9535197; PID:9535198
 C:Superfamily: collagen alpha 1(IV) chain

Query Match 90.9%; Score 1218.5; DB 2; Length 246;
 Best Local Similarity 90.2%; Pred. No. 2e-103;
 Matches 221; Conservative 12; Mismatches 11; Indels 1; Gaps 1;

QY 1 GLKRGDSSPATWTRGFEVTRHSQTTATPSCPEGVPLYSGFSEFLVQGNRAHQ 59
 2 GLKRGDSSPATWTRGFEVTRHSQTTATPSCPEGVPLYSGFSEFLVQGNRAHQ 61
 QY 60 DIGTIGSCLQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALMNMNAPITGRALPEYI 119
 62 DIGTIGSCLQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALMNMNAPITGRALPEYI 121
 QY 120 SRCYVCEGPAIAIVHSQTTDIPCPCHGWSLWKGFSTLMTFSASSEGQALASPGSCL 179
 122 SRCYVCEGPAIAIVHSQTTDIPCPCHGWSLWKGFSTLMTFSASSEGQALASPGSCL 181
 QY 180 EFRASPLECHRGRCNCYNSYSFWLNLNPERFRKPISTYKAGLEKIIISRCQY 239
 182 EFRASPLECHRGRCNCYNSYSFWLNLNPERFRKPISTYKAGLEKIIISRCQY 241
 QY 240 MKRH 244
 242 MKRH 246

RESULT 3
 collagen alpha 3(IV) chain - bovine (fragment)
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
 C:Accession: A39024; S20672; S17802; A35167; C39419; S13747; S20815
 R:Morrison, K.E.; Germino, G.G.; Reeders, S.T.
 J. Biol. Chem. 266, 34-39, 1991
 A:Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding t
 A:Reference number: A39024; MUID:91093146
 A:Accession: A39024
 A:Molecule type: mRNA
 A:Residues: 1-471 <MOR>
 A:Cross-references: EMBL:M63139; NID:9162886; PIDN:AA62708.1; PID:9162887
 R:Butkowski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.
 J. Biol. Chem. 262, 7874-7877, 1987
 A:Title: Localization of the Goodpasture epitope to a novel chain of basement membran
 A:Reference number: S18432; MUID:87222419
 A:Accession: S20672
 A:Molecule type: protein
 A:Residues: 227-228, 'X', 230-244 <BUT>
 R:Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.
 J. Biol. Chem. 263, 13374-13380, 1988
 A:Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collag
 A:Reference number: S17802; MUID:88330844
 A:Accession: S17802
 A:Molecule type: protein
 A:Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>
 R:Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.
 J. Biol. Chem. 265, 5466-5469, 1990
 A:Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of t
 A:Reference number: A35167; MUID:90202779
 A:Accession: A35167
 A:Molecule type: protein
 A:Residues: 236-238 <GUN>
 R:Gunwar, S.; Ballester, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.;
 J. Biol. Chem. 266, 15318-15324, 1991
 A:Title: Glomerular basement membrane. Identification of dimeric subunits of the nonc

R:Bratzel, D.; Oberdaewer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann, Eur. J. Biochem. 168, 529-536, 1987

A:Title: Completion of the amino acid sequence of the alpha1 chain of human basement A:Reference number: S00048; MUID:88029471

A:Accession: S00048

A:Molecule type: mRNA

A:Residues: 1-318, 'A', 320-944 <BRA1>

A:Cross-references: EMBL:X05561; NID:g30066; PIDN:CAA29075.1; PID:g30067

A:Accession: S25826

A:Molecule type: protein

A:Residues: 271-318, 'A', 320-554 <BRA2>

R:Glanville, R.W.; Qian, R.O.; Siebold, B.; Ristell, J.; Kuehn, K.

Eur. J. Biochem. 152, 213-219, 1985

A:Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (A:Reference number: A23115; MUID:86004708

A:Accession: A23115

A:Molecule type: protein

A:Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GLA>

A:Experimental source: placenta

A:Note: The amino end of the mature form is blocked

R:Solminen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggvason, K.

FEBS Lett. 225, 188-194, 1987

A:Title: Complete primary structure of the alpha(1)-chain of human basement membrane A:Reference number: S00207; MUID:88083584

A:Accession: S00207

A:Molecule type: mRNA

A:Residues: 244-530 <SOI3>

A:Cross-references: EMBL:Y00706; NID:g29548; PIDN:CAA6698.1; PID:g29549

R:Ehle, J.A.; Golik, R.; Mann, K.; Kuehn, K.

EMBO J. 12, 4795-4802, 1993

A:Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collagen A:Reference number: S39614; MUID:94038963

A:Accession: S39614

A:Molecule type: protein

A:Residues: 371-554 <EBD>

R:Babel, W.; Glanville, R.W.

Eur. J. Biochem. 143, 545-556, 1984

A:Title: Structure of human-basement-membrane (type IV) collagen. Complete amino-acid A:Reference number: A02863; MUID:85003629

A:Accession: A02863

A:Molecule type: protein

A:Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 9 A:Experimental source: placenta

R:Glanville, R.W.; Rauter, A.

Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981

A:Title: Peptic fragments of human placental basement-membrane collagens showing intenc A:Reference number: S16908; MUID:82005835

A:Accession: A58517

A:Molecule type: protein

A:Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553, 1389-1405, 'XX', 1408-1409, 'X', 1411 R:MacWright, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.

Biochemistry 22, 4940-4948, 1983

A:Title: Isolation and characterization of pepsin-solubilized human basement membrane A:Reference number: S16910; MUID:84053346

A:Accession: S16910

A:Molecule type: protein

A:Residues: 534-537, 'G', 539, 'G', 541-542, 'G', 544-549, 939-940, 'W', 942-944, 'V', 946, 'X', 9, 9 A:Experimental source: placenta

R:PhilaJammi, T.; Tryggvason, K.; Myers, J.C.; Kurkinen, M.; Lebo, R.; Cheung, M.C.

J. Biol. Chem. 260, 7681-7687, 1985

A:Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen A:Reference number: S01466; MUID:85207819

A:Accession: S01466

A:Molecule type: mRNA

A:Residues: 1256-1669 <PIH>

A:Cross-references: EMBL:M10940; NID:g180421; PIDN:AAA52006.1; PID:g180424

R:Brinker, J.M.; Gudes, L.J.; Loidl, H.R.; Wang, S.Y.; Rosenbloom, J.; Kefalides, N.A.; Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985

A:Title: Restricted homology between human alpha-1 type IV and other procollagen chain A:Reference number: S16879; MUID:85216555

A:Accession: S16879

A:Molecule type: mRNA

A:Residues: 1259-1669 <BRI>

FEBS Lett. 227, 5-8, 1988
A:Title: CDNA clones completing the nucleotide and derived amino acid sequence of the al
A:Reference number: S01454; MUID:88112221
A:Accession: S01454
A:Molecule type: mRNA
A:Residues: 1-185, 'L', 187-318, 'S', 320-368, 'L', 370-402, 'F', 404-480, 'L', 482-492, 'H', 494-71
A:Cross-references: EMBL:X06777
R:Killem, P.D.; Burdello, P.; Sakurai, Y.; Yamada, Y.
J. Biol. Chem. 263, 8706-8709, 1988
A:Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen cha
A:Reference number: A28066; MUID:88243724
A:Accession: A28066
A:Molecule type: mRNA
A:Residues: 1-129 <K13>
A:Cross-references: EMBL:J03758; NID:9192669; PIDN:AAA37439.1; PID:9192670
R:Oberbauer, I.; Laurent, M.; Schwartz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss,
Eur. J. Biochem. 147, 217-224, 1985
A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
A:Reference number: A02864; MUID:85127033
A:Accession: A02864
A:Molecule type: mRNA
A:Residues: 1276-1669 <OB>
A:Cross-references: EMBL:X02201; NID:950233; PIDN:CAA26132.1; PID:91333876
R:Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G.
Gene 43, 301-304, 1986
A:Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeox
A:Reference number: A25636; MUID:86301886
A:Accession: A25636
A:Molecule type: mRNA
A:Residues: 1149-1396, 'S', 1398-1424 <NAT>
A:Cross-references: EMBL:M14042; NID:9192286; PIDN:AAA37342.1; PID:9192287
A:Note: the authors translated the codon CAG for residue 1374 as Arg
R:Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihla
J. Biol. Chem. 262, 8496-8499, 1987
A:Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)
A:Reference number: A94680; MUID:87250460
A:Accession: A29301
A:Molecule type: mRNA
A:Residues: 1441-1669 <KUR>
A:Cross-references: EMBL:M15832; NID:9192282; PIDN:AAA37340.1; PID:9387115
R:Killem, P.D.; Burdello, P.D.; Martin, G.R.; Yamada, Y.
J. Biol. Chem. 263, 12310-12314, 1988
A:Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequenc
A:Reference number: S19079; MUID:88315019
A:Accession: S19079
A:Molecule type: DNA
A:Residues: 1-28 <K1>
A:Cross-references: EMBL:J03944; NID:9192673; PIDN:AAA37442.1; PID:9466503
R:Hayes, P.; Wood, L.; Thieriault, N.; Kurkinen, M.; Vogeli, G.
Mol. Chem. 263, 19274-19277, 1988
A:Title: Head-to-head arrangement of murine type IV collagen genes.
A:Reference number: A92702; MUID:89066738
A:Accession: A32003
A:Molecule type: DNA
A:Residues: 1-28 <KAV>
A:Cross-references: EMBL:J04448; NID:9192666; PIDN:AAA37437.1; PID:9450449
R:Burdello, P.D.; Martin, G.R.; Yamada, Y.
Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988
A:Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional prom
A:Reference number: A94220; MUID:89071759
A:Accession: A31766
A:Molecule type: DNA
A:Residues: 1-28 <BUR>
A:Cross-references: EMBL:M23333; NID:9340878; PIDN:AAA51625.1; PID:9535668
R:Sakurai, Y.; Sullivan, M.; Yamada, Y.
J. Biol. Chem. 261, 6654-6657, 1986
A:Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes
A:Reference number: S19094; MUID:86196099
A:Accession: S19094
A:Molecule type: DNA
A:Residues: 1110-1135, 1189-1316, 1342-1383, 1418-1487 <SAK>
A:Cross-references: EMBL:M13027
R:Schuppan, D.; Timpl, R.; Glatville, R.W.

FEBS Lett. 115, 297-300, 1980
A:Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane
A:Reference number: S16909; MUID:80246483
A:Accession: S16909
A:Molecule type: protein
A:Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957, 1213-1228, 'X', 1230-1234, 'P', 1236-
R:Schuppan, D.; Glatville, R.W.; Timpl, R.
Eur. J. Biochem. 123, 505-512, 1982
A:Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial a
A:Reference number: A25991; MUID:82186723
A:Accession: A25991
A:Molecule type: protein
A:Residues: 940-946, 'X', 948-949, 'X', 951-955, 'X', 957-964, 'X', 966-991, 'X', 993-1003, 'X',
61, 'X', 1063-1065, 'X', 1067-1080, 'X', 1082-1083, 'X', 1085-1106, 'X', 1108-1115, 'DE', 1118-11
A:Accession: B25991
A:Molecule type: protein
A:Residues: 1173-1181, 'X', 1183-1184, 'X', 1186-1187, 'X', 1189-1205, 'O', 1207, 'XE', 1210-12
3, 'SP', 1266, 'IR', 1269, 'SK', 1272, 'DW', 1275, 'L', 1277-1282, 1316-1318, 'X', 1320-1327, 'X', 1
R:Weber, S.; Engel, J.; Wiedemann, H.; Glatville, R.W.; Timpl, R.
Eur. J. Biochem. 139, 401-410, 1984
A:Title: Subunit structure and assembly of the globular domain of basement-membrane c
A:Reference number: S17801; MUID:84132058
A:Accession: S17801
A:Molecule type: protein
A:Residues: 1435-1443 <WEB>
A:Genetics: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3
A:Note: The list of introns may be incomplete
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane; cell binding; coiled coil; duplication; extracellular
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-1669/Product: collagen alpha 1(IV) chain #status predicted <MAT>
F:28-162/Domain: 7S <7SD>
F:163-1460/Domain: collagenous, triple helix <COL>
F:159-159/Region: cell attachment (R-G-D) motif
F:781-783/Region: cell attachment (R-G-D) motif
F:917-919/Region: cell attachment (R-G-D) motif
F:968-970/Region: cell attachment (R-G-D) motif
F:1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F:1441-1552/Region: duplication
F:1553-1669/Region: duplication
F:31,36,39,41,434,467,470/Disulfide bonds: interchain #status predicted
F:126/Binding site: carbohydrate (asn) (covalent) #status predicted
F:971,974,977,986,989,1001,1007,1019,1022,1031,1037,1040,1055,1060,1063,1075,1078,109
92,1298,1310,1313,1322,1337,1346,1349,1422,1425,1431,1437,1440/Modified site: hydroxy
F:1214,1424/Modified site: 4-hydroxyproline (Pro) #status experimental
F:1304/Modified site: 5-hydroxylysine (Lys) #status experimental
F:1505-1511,1516-1622/Disulfide bonds: #status predicted

Query Match 71.0%; Score 951; DB 1; Length 1669;
Best Local Similarity 68.6%; Pred. No. 3.2e-78;
Matches 166; Conservative 34; Mismatches 40; Indels 2; Gaps 1;

1 GLKGRGDSGSPATWTTTGCVFTRHSQTTAIPSCBEGTVPLSGSFVLVGNORAHGOD 60
1429 GLPGMGPPGTPS--VDHGFVLTTRHSQTTDPLCPGTVILHGVSLVQDNERAHGOD 1486
DB 1429 GLPGMGPPGTPS--VDHGFVLTTRHSQTTDPLCPGTVILHGVSLVQDNERAHGOD 1486
61 LTTTSCLORFETMPDFLPCNVNDVCFASRNDYSYWLSTPALPMNMAPTGRALEPYTS 120
DB 1487 LCTASCLRKPFSTMPFLCINNVNCFASRNDYSYWLSTPALPMNMAPTGRALEPYTS 1546
DB 1487 LCTASCLRKPFSTMPFLCINNVNCFASRNDYSYWLSTPALPMNMAPTGRALEPYTS 1546
121 RCTVCEGPAIAIVHSQTTDPPCPHGWISLWKGFSTFMTSAGSBGTQALASPGSCLE 180
DB 1547 RCVAWEAPAMWAVHSQTTDPPCPHGWISLWKGFSTFMTSAGSBGTQALASPGSCLE 1606
DB 1547 RCVAWEAPAMWAVHSQTTDPPCPHGWISLWKGFSTFMTSAGSBGTQALASPGSCLE 1606
181 EFRASPFLECHRGICNYNSYNSFVWLASINPERMRKRPSTVAGAELEKTIISCOVCM 240
DB 1607 EFRASPFLECHRGICNYNSYNSFVWLASINPERMRKRPSTVAGAELEKTIISCOVCM 240
DB 1607 EFRASPFLECHRGICNYNSYNSFVWLASINPERMRKRPSTVAGAELEKTIISCOVCM 240
OY 241 KR 242
OY 241 KR 242
DB 1667 RR 1668

RESULT
7
S22917
collagen alpha 5(IV) chain precursor, renal splice form - human
N:Alternate names: procollagen alpha 5(IV) chain
N:Contents: collagen alpha 5(IV) chain precursor, leukocyte splice form
C:Species: Homo sapiens (man)
C:Date: 30-Sep-1993 #sequence, revision 27-Feb-1997 #text change 21-Jul-2000
C:Accession: S22917; A53079; A37122; A34850; S18850; I56971; I76598; A35
R:Zhou, J.; Hertig, J.M.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 267, 12475-12481, 1992
A:Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identifi
n Alport syndrome patient.
A:Reference number: S22917; MUID:92316923
A:Accession: S22917
A:Molecule type: mRNA
A:Residues: 1-967 <CHO>
A:Cross-references: GB:M90464; NID:q180826; PIDN:AAA52046.1; PID:q553234
R:Zhou, J.; Leinonen, A.; Tryggvason, K.
J. Biol. Chem. 269, 6608-6614, 1994
A:Title: Structure of the human type IV collagen COL4A5 gene.
A:Reference number: A54365; MUID:94165049
A:Accession: A54365
A:Molecule type: DNA
A:Residues: 1-922 <RH2>
A:Cross-references: GB:U04470; NID:q463378; GB:U04520; NID:q463428; PIDN:AAC27816.1; PID
R:Zhou, J.; Mochizuki, T.; Smeets, H.; Antignac, C.; Laurila, P.; de Paep, A.; Tryggvason
Science 261, 1167-1169, 1993
A:Title: Deletion of the paired alpha5(IV) and alpha6(IV) collagen genes in inherited sm
A:Reference number: A57079; MUID:93361972
A:Accession: A57079
A:Molecule type: DNA
A:Residues: 1-27 <ZH4>
A:Cross-references: GB:J27153; NID:q587203; PIDN:CAA5512.1; PID:q587204
R:Phlajaniemi, T.; Pohjolahti, E.R.; Myers, J.C.
J. Biol. Chem. 265, 13758-13766, 1990
A:Title: Complete primary structure of the triple-helical region and the carboxyl-termin
A:Reference number: A37122; MUID:90337990
A:Accession: A37122
A:Molecule type: mRNA
A:Residues: 84-439, 'GS', 442-624, 'LALO', 629-666, 'RR', 669-887, 'R', 889-1264, 1271-1691 <PIH>
A:Cross-references: GB:J05558; EMBL:M58526; NID:q1314209
A:Note: submitted to the EMBL Data Library, February 1991
A:Note: the authors translated the codon GCC for residue 115 as Val
R:Renier, A.; Seril, M.; Myers, J.C.; Phlajaniemi, T.; Massella, L.; Rizzoni, G.; De Ma
Hum. Mol. Genet. 1, 127-129, 1992
A:Title: De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in
A:Reference number: I54317; MUID:93244772
A:Accession: I54317
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 313-324, 'E', 326-330 <REN>
A:Cross-references: GB:S59334; NID:q29946; PIDN:AAD13909.1; PID:q4261609
R:Hostikka, S.L.; Eddy, R.L.; Byers, M.G.; Hoerlyhae, M.; Shows, T.B.; Tryggvason, K.
Proc. Natl. Acad. Sci. U.S.A. 87, 1606-1610, 1990
A:Title: Identification of a distinct type IV collagen alpha chain with restricted kidn
A:Reference number: A34850; MUID:90160375
A:Accession: A34850
A:Molecule type: mRNA
A:Residues: 914-1264, 1271-1691 <HOS>
A:Cross-references: EMBL:M31115; NID:q180824; PIDN:AAA52045.1; PID:q180825
R:Zhou, J.; Hostikka, S.L.; Chow, L.T.; Tryggvason, K.
Genomics 9, 1-9, 1991
A:Title: Characterization of the 3' half of the human type IV collagen alpha-5 gene that
A:Reference number: A37969; MUID:91169491
A:Accession: S18850
A:Molecule type: DNA
A:Residues: 924-1264, 1271-1691 <ZH3>
A:Cross-references: EMBL:M63455; EMBL:M63456; EMBL:M63457; EMBL:M63458; EMBL:M63459; EMB
8; EMBL:M63470; EMBL:M63471; EMBL:M63472; EMBL:M63473; NID:q177922; PIDN:AAA51558.1; PID
R:Guo, C.; Van Damme, B.; Van Damme-Lombaerts, R.; Van den Berghe, H.; Cassiman, J.J.; M
Kidney Int. 44, 1316-1321, 1993

A:Title: Differential splicing of COL4A5 mRNA in kidney and white blood cells: a comp
A:Reference number: I56971; MUID:94133540
A:Accession: I56971
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1258-1276 <GUO1>
A:Cross-references: GB:S69168; NID:q545095; PIDN:AAC60612.1; PID:q545096
A:Note: kidney splice form
A:Accession: I76598
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1284-1291, 'FTGLVLAIV' <GUO2>
A:Cross-references: GB:S69169; NID:q545097; PIDN:AAC60613.1; PID:q545098
A:Note: frameshift mutation in patient with Alport syndrome
R:Myers, J.C.; Jones, T.A.; Pohjolahti, E.R.; Kadri, A.S.; Goddard, A.D.; Sheer, D.;
Am. J. Hum. Genet. 46, 1024-1033, 1990
A:Title: Molecular cloning of alpha5(IV) collagen and assignment of the gene to the r
A:Reference number: A35335; MUID:90252791
A:Accession: A35335
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1448-1477 <MYE>
R:Nakazato, H.; Hattori, S.; Ushijima, T.; Matsura, T.; Kaitabashi, Y.; Takada, T.;
Kidney Int. 46, 1307-1314, 1994
A:Title: Mutations in the COL4A5 gene in Alport syndrome: a possible mutation in prim
A:Reference number: I56975; MUID:95156893
A:Accession: I56975
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1595-1602 <NAK>
A:Cross-references: GB:S75903; NID:q913882; PIDN:AAB33374.1; PID:q913883
A:Note: premature termination mutation from a patient with Alport syndrome: one other
R:Lemmink, H.H.; Schroeder, C.H.; Brunner, H.G.; Nelen, M.R.; Zhou, J.; Tryggvason, K
Genomics 17, 485-489, 1993
A:Title: Identification of four novel mutations in the COL4A5 gene of patients with A
A:Reference number: I54188; MUID:94010948
A:Accession: I54188
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1604-1607, 'VHDAYKC' <LEM>
A:Cross-references: GB:S6567; NID:q425563; PIDN:AAD13967.1; PID:q4261667
A:Note: frameshift mutation from a patient with Alport syndrome: five other mutations
C:Comment: Prolines and lysines at the third position of the tripeptide repeating uni
ed and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB:COL4A5; AYS
A:Cross-references: GDB:120596; OMIM:303630
A:Map position: Xq22-Xq22
A:Insertions: 27/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 2
/3; 799/1; 837/1; 893/1; 923/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1185/1
A:Note: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands w
C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha
moy trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeri
er associations in the interrupted helical domain (with disulfide and desmosine cross
C:Function:
A:Description: minor structural component of extracellular basement membrane
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: Alport syndrome; basement membrane; coiled coll.; extracellular matrix; gl
F:1-26/Domain: signal sequence #status predicted <Sig>
F:27-1691/Product: collagen alpha 5(IV) chain, renal splice form #status predicted <M
F:27-1264, 1271-1691/Product: collagen alpha 5(IV) chain, leukocyte splice form #statu
F:27-41/Domain: amino-terminal nonhelical, NC2 #status predicted <NC2>
F:1463-1691/Region: interrupted helical
F:1463-1573/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F:1583-1667/Domain: collagen IV carboxyl-terminal repeat <CR1>
F:29,32,38,40,124,451,481,484/Disulfide bonds: interchain #status predicted
F:125/Binding site: carbonyl (Asn) (covalent) #status predicted
F:1483-1570, 1515-1573/Disulfide bonds: (or 1482-1573, 1515-1570) #status predicted
F:1527-1533, 1638-1644/Disulfide bonds: #status predicted
F:1592-1664, 1626-1667/Disulfide bonds: (or 1592-1687, 1626-1684) #status predicted

F;34-134/Domain: collagen IV carboxyl-terminal repeat <CT1>

Query March 65.1%; Score 872; DB 2; Length 220;
 Best Local Similarity 99.4%; Pred. No. 5,4e-72;
 Matches 158; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	1	GLKGRGDSGSPATWTRTGRFVETRH	SQTALP	SCPEGVPLYSGFSEFLVQGN	RAHGQD	60
DB	10	GLKGRGDSGSPATWTRTGRFVETRH	SQTALP	SCPEGVPLYSGFSEFLVQGN	RAHGQD	69
QY	61	LGTLASCLQRTTFMFLFCN	VNDFAS	RNDYSWTSW	TPALMNNAP	IGRALPEYIS 120
DB	70	LGTLASCLQRTTFMFLFCN	VNDFAS	RNDYSWTSW	TPALMNNAP	IGRALPEYIS 129
QY	121	RCTVCGEPALIAVAHSQ	TTDIPPC	PHGWSILMKGS	SEFTM	159
DB	130	RCTVCGEPALIAVAHSQ	TTDIPPC	PHGWSILMKGS	SEFTM	168

[illegible]

A:Reference number: 513651
A:Accession: S13651
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-139 / GFGMPDPLAGPGSGNGNPNRGLSGPGEGVNSGCGKVGKSGSGRGVGLT', 209-281, 'P
15 / D', 817-1260, 'P', 1262-1515 <CU1>
A:Cross-references: EMBL:X56979
R:Guo, X.; Kramer, J.M.
J. Biol. Chem. 264, 17574-17582, 1989
A:Title: The two Cenorhabditis elegans basement membrane (type IV) collagen genes are 1
A:Reference number: A34476; MUID:90008929
A:Accession: B34476
A:Molecule type: DNA
A:Residues: 1433-1499, 'Q', 1501-1707, 'P', 1709-1744 <CU2>
A:Cross-references: EMBL:J05067; NID:g156255; PIDN:AAB59179.1; PID:g156256
C:Genetics:
A:Gene: clb-2; emb-9
A:Map position: 3
A:Intons: 23/2; 79/1; 152/2; 288/1; 329/3; 391/1; 575/3; 660/3; 741/3; 1028/3; 1453/1;
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane, triple helix #status predicted <CDL>
F:43-1515/Domain: collagenous, triple helix #status predicted <CDL>
F:33-95/Region: cell attachment (R-G-D) motif
F:1053-1055/Region: cell attachment (R-G-D) motif
F:1396-1398/Region: cell attachment (R-G-D) motif
F:1516-1744/Domain: carboxyl-terminal nonhelical, Ncl #status predicted <NCL>
F:1516-1627, 1628-1744/Region: duplication
F:1580-1586, 1691-1697/Disulfide bonds: #status predicted

A:Accession: A45407
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1752 <EXP>
A>Note: sequence extracted from NCBI backbone (NCBI:P126841)
R:Wessel, G.M.; Etkin, M.; Benson, S.
Dev. Biol. 148, 261-272, 1991
A:Title: Primary mesenchyme cells of the sea urchin embryo require an autonomously produced reference number: A43903; MUID:92038439
A:Accession: A43903
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 'P', 633-1537, 'G', <MES>
A:Cross-references: GB:S64572; NID:g238616; PIDN:AA020270.1; PID:g238617
A>Note: sequence extracted from NCBI backbone (NCBI:64572; NCBI:P:64573)
R:Venkatesan, M.; De Pablo, F.; Vogeli, G.; Simpson, R.T.
Proc. Natl. Acad. Sci. U.S.A. 83, 3351-3355, 1986
A:Title: Structure and developmentally regulated expression of a Strongylocentrotus purpuratus reference number: A23940; MUID:86205894
A:Accession: A23940
A:Molecule type: DNA
A:Residues: 742-812 <VEN>
A:Cross-references: EMBL:M13206
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: coll: coll; extracellular matrix; glycoprotein; trimer; triple helix
F:29-161/Domain: amino-terminal nonhelical, 75 <7SD>
F:162-1523/Region: interrupted helical
F:1524-1752/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F:1534-1634/Domain: collagen IV carboxyl-terminal repeat <CT1>
F:1644-1748/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:129/Modified site: allysine (lys) #status predicted

Query Match 62.1%; Score 832; DB 2; Length 1752;
Best Local Similarity 59.3%; Pred. No. 2.2e-67;
Matches 144; Conservative 37; Mismatches 60; Indels 2; Gaps 1;

QY 1 GLKGRGDSGSPATWTTGCVFTRHSQTTAIPSCBPGVPLYSGFLEVCNORAHGOD 60
DB 1512 GPPGKGGAGTIPG--SSSGFTTRHSQTTIPQCPGTAKMNHGSLLEPQENRKHGOD 1569
QY 61 LGTSGCLQRTTTPFLPCNVNDVNCNPFASRNDYSYWLSTPALPMNMAPITGRALEPYIS 120
DB 1570 LGKPGSCLRKSTMPFLPCNINNVCHVASRNDYSYWLSTTEPMNMAPIRGQLOPFTIS 1629
QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGMSLWKGSFTIMFTSASGEGTQALASPGSCLE 180
DB 1630 RCVCCEAPAOVLTVHSQTVNIPDCPDRCVLMIGYSFMNHTGPGGEGSQMLSSPGSCLE 1689
181 EFRASPFLBCHG-RGTCNYSNSYSFWLASLNERMFRKPISTVAGLEKTIISRCQVCM 240
DB 1690 DFRASPFLBCHGDKCNVYATITTFWSSITGNQFTMPQSETLAKGSLRTVRSRCVCL 1749
QY 241 KKR 243
DB 1750 RNO 1752
RESULT 14
S16366
collagen alpha 2(IV) chain precursor - pig roundworm
C:Species: Ascaris suum (pig roundworm)
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C:Accession: S16366
R:Pettilt, J.; Kingston, I.B.
J. Biol. Chem. 266, 16149-16156, 1991
A:Title: The complete primary structure of a nematode alpha-2(IV) collagen and the partial reference number: S16366; MUID:91340768
A:Accession: S16366
A:Molecule type: mRNA
A:Residues: 1-1763 <CBI>
A:Cross-references: GB:M67507; NID:g159648; PIDN:AAA18014.1; PID:g159649
C:Genetics:

A:Introns: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1; 963/3
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disul
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-1763/Product: collagen alpha 2(IV) chain #status predicted <MAT>
F:27-42/Domain: non-collagenous NH1 #status predicted <NH1>
F:43-1529/Domain: collagenous #status predicted <COL>
F:129-199/Region: cell attachment (R-G-D) motif
F:1530-1763/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F:1530-1638/Domain: repeat NC1 #status predicted <NC12>
F:1639-1763/Domain: repeat NC1 #status predicted <NC12>
F:31.34.39.41.536.539/Disulfide bonds: interchain #status predicted
F:126/Binding site: carbohydrate (asn) (covalent) #status predicted
F:1593-1599, 1702-1709/Disulfide bonds: #status predicted

Query Match 59.4%; Score 796.5; DB 2; Length 1763;
Best Local Similarity 57.9%; Pred. No. 3.8e-64;
Matches 140; Conservative 41; Mismatches 58; Indels 3; Gaps 2;

QY 1 GLKGRGDSGSPATWTTGCVFTRHSQTTAIPSCBPGVPLYSGFLEVCNORAHGOD 60
DB 1515 GLPGSPGPGPGPPGPKGDFLLVKHSQTSSEVPQCPGMWKLMDGYSLYIEGNEKSHND 1574
QY 61 LGTSGCLQRTTTPFLPCNVNDVNCNPFASRNDYSYWLSTPALPMNMAPITGRALEPYIS 120
DB 1575 LGHASSCLSRFTMPFLPCNVNDVNCNPFASRNDYSYWLSTTA--PIPMHVSAGGLEPYIS 1632
QY 121 RCTVCEGPAIAIAVHSQTTDIPPCPHGMSLWKGSFTIMFTSASGEGTQALASPGSCLE 180
DB 1633 RCVCCEAPANVIAVHSQTIQIPNCNMGNSLWIGYSFAMHTGAGAGEGGQSLSSPGSCLE 1692
QY 181 EFRASPFLBCHG-RGTCNYSNSYSFWLASLNERMFRKPISTVAGLEKTIISRCQVCM 239
DB 1693 DFRASPFLBCHGRCGTCNYSNSYSFWLASLNERMFRKPISTVAGLEKTIISRCQVCM 1752
QY 240 MK 241
DB 1753 IR 1754

RESULT 15
T29350
hypothetical protein F01G12.5a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T29350
R:Wu, X.; Le, T.T.
submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid F01G12.
A:Reference number: Z20611
A:Accession: T29350
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1758 <MUX>
A:Cross-references: EMBL:U53342; PIDN:AAA96216.1; GSPDB:GN00028; CESP:F01G12.5a
A:Experimental source: strain Bristol N2; clone F01G12
C:Genetics:
A:Gene: CESP:F01G12.5a
A:Map position: X
A:Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 264/3; 303/3; 358/3; 449/2; 73
C:Superfamily: collagen alpha 1(IV) chain

Query Match 58.5%; Score 783.5; DB 2; Length 1758;
Best Local Similarity 55.3%; Pred. No. 5.7e-63;
Matches 142; Conservative 41; Mismatches 59; Indels 15; Gaps 4;

QY 1 GLKGRGDSGSP-----ATWTTGCVFTRHSQTTAIPSCBPGVPLYSGFLEVCNORAHGOD 51
DB 1504 GLDGPPGGGAGALGPAGPACPAYRDGFLVKHSQTEVPRCPGQRLMDGYSLYIE 1563
QY 52 GNORAHGODLGLGSCLOQRTTTPFLPCNVNDVNCNPFASRNDYSYWLSTPALPMNMAPIT 111

```
Db 1564 GNEKSHNODLGHAGSCIQRSTMPLECDENNVCNYSRNDKSYWLSTSEAI--MMPVN 1621
QY 112 GRALEPIISRCTYCEGAIATAVHSQTTDIPCPHGWIISLKGFSFIMFTSAGSEGTGA 171
Db 1622 EREIEPIISRCAYCEAPANTIAVHSQTTQIPNCPAGWSSIMIGSFAMHTGAGAGGGOS 1681
QY 172 LASPGSCLFEFRASPFECHG-ROTCNYSNSYSFWLASINPERMFRKPIPSTVKAGELE 230
Db 1682 LSSPGSCLDEPRATPFIECHGARGSCHYFANKFSFWLTTIDNDSERFVPESSQTLKSGNLR 1741
QY 231 KIISRCOVCMKK--RH 244
Db 1742 TRVSRCCQCVKSTDGRH 1758
```

Search completed: April 13, 2001, 15:42:38
Job time: 219 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2001, 15:41:57 ; Search time 31.4 Seconds
(without alignments)
43.439 Million cell updates/sec

Title: US-09-543-371-10_COPY_54_124
Perfect score: 390
Sequence: 1 ORAHGDLGTLSGCLQRFRTT.....MNAPIITGRALEPYISRCTV 71

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/2/1aa/PTUS.COMB.pep:*
- 6: /cgn2_6/ptodata/2/1aa/Dackfilest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	390	100.0	218	2	US-08-399-889-25
2	390	100.0	218	3	US-09-167-364-25
3	371	95.1	471	2	US-08-399-889-24
4	371	95.1	471	3	US-09-167-364-24
5	144	36.9	1694	1	US-08-494-168-2
6	63	16.2	628	2	US-07-952-853-22
7	63	16.2	628	2	US-08-914-848-22
8	60	15.4	186	2	US-08-766-551-3
9	57	14.6	397	1	US-08-098-141-2
10	55	14.1	3898	2	US-08-876-991-2
11	55	14.1	3898	2	US-09-059-853-2
12	54.5	14.0	520	3	US-08-949-386-28
13	54.5	14.0	520	3	US-08-450-562-28
14	54	13.8	472	4	US-08-976-255-17
15	53.5	13.7	574	4	US-08-895-590-7
16	53	13.6	612	3	US-08-482-677-2
17	53	13.6	612	3	US-08-762-500-79
18	52.5	13.5	364	1	US-08-318-831-6
19	52	13.3	275	4	US-08-976-235-16
20	52	13.3	282	3	US-08-812-586-49
21	51.5	13.2	1333	4	US-09-356-952-2
22	51	13.1	109	2	US-08-467-822-24
23	51	13.1	277	1	US-08-118-270-62
24	51	13.1	277	5	PCT-US93-08528-62
25	51	13.1	3898	4	US-08-750-717-2
26	50.5	12.9	354	2	US-08-700-013B-9
27	50.5	12.9	374	2	US-08-820-170A-25

28	50.5	12.9	374	3	US-09-055-699-25	Sequence 25, Appl
29	50.5	12.9	374	4	US-09-273-565-25	Sequence 25, Appl
30	50.5	12.9	374	5	PCT-US95-06385-2	Sequence 2, Appl
31	50.5	12.9	515	2	US-09-073-362-3	Sequence 3, Appl
32	50.5	12.9	515	2	US-09-243-920-3	Sequence 3, Appl
33	50.5	12.9	797	2	US-08-700-013B-21	Sequence 21, Appl
34	50	12.8	1319	2	US-08-290-731C-2	Sequence 6, Appl
35	49.5	12.7	245	2	US-08-825-781-1	Sequence 1, Appl
36	49.5	12.7	511	2	US-09-073-362-1	Sequence 1, Appl
37	49.5	12.7	511	2	US-09-243-920-1	Sequence 1, Appl
38	49.5	12.7	511	2	US-09-022-983-5	Sequence 5, Appl
39	49.5	12.7	740	4	US-08-976-255-18	Sequence 18, Appl
40	49	12.6	178	4	US-08-415-751-8	Sequence 8, Appl
41	49	12.6	361	1	US-08-845-539-6	Sequence 6, Appl
42	48.5	12.4	179	2	US-08-700-013B-6	Sequence 6, Appl
43	48.5	12.4	405	2	US-08-700-013B-6	Sequence 6, Appl
44	48.5	12.4	797	2	US-08-700-013B-19	Sequence 19, Appl
45	48.5	12.4	799	2	US-08-700-013B-27	Sequence 27, Appl

ALIGNMENTS

```
RESULT 1
US-08-399-889-25
; Sequence 25, Application US/08399889B
; Patent No. 5973120
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263A
; CURRENT APPLICATION NUMBER: US/08/399, 889B
; EARLIER FILING DATE: 1995-03-07
; EARLIER APPLICATION NUMBER: 07/621091
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 25
; LENGTH: 218
; TYPE: PRT
; ORGANISM: Human
US-08-399-889-25

Query Match      100.0%; Score 390; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 2.5e-43;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORAHGDLGTLSGCLQRFRTTMPFLFCNVNDVCFASRNDYSYWLSTPALPMNAPIITGR 60
    |||||||
DB 28 ORAHGDLGTLSGCLQRFRTTMPFLFCNVNDVCFASRNDYSYWLSTPALPMNAPIITGR 87
    |||||||

QY 61 ALEPYISRCTV 71
    |||||||
DB 88 ALEPYISRCTV 98

RESULT 2
US-09-167-364-25
; Sequence 25, Application US/09167364
; Patent No. 6007980
; GENERAL INFORMATION:
; APPLICANT: Reeders, Stephen T
; APPLICANT: Morrison, Karen E
; APPLICANT: Hudson, Billy G
; TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
; FILE REFERENCE: 951263B
; CURRENT APPLICATION NUMBER: US/09/167, 364
; EARLIER FILING DATE: 1998-10-07
; EARLIER APPLICATION NUMBER: 08/399889
; EARLIER FILING DATE: 1995-03-07
```

NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 25
LENGTH: 218
TYPE: PRT
ORGANISM: Human
US-09-167-364-25

Query Match
Best Local Similarity 100.0%; Score 390; DB 3; Length 218;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORAHGODLGTGSCLOPFTMPFLFCNVNDVCFNFSRNDYSYWLSTPALMPMNPATIGR 60
DB 28 ORAHGODLGTGSCLOPFTMPFLFCNVNDVCFNFSRNDYSYWLSTPALMPMNPATIGR 87

QY 61 ALEPIISRCTV 71
88 ALEPIISRCTV 98

RESULT 3
US-08-399-889-24

Sequence 24, Application US/08399889B
Patent No. 5973120
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T
APPLICANT: Morrison, Karen E
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 951263A
CURRENT APPLICATION NUMBER: US/08/399, 889B
CURRENT FILING DATE: 1995-03-07
EARLIER APPLICATION NUMBER: 07/621091
EARLIER FILING DATE: 1990-11-30
NUMBER OF SEQ ID NOS: 25
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 24
LENGTH: 471
TYPE: PRT
ORGANISM: Calf
US-08-399-889-24

Query Match
Best Local Similarity 95.1%; Score 371; DB 2; Length 471;
Matches 65; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORAHGODLGTGSCLOPFTMPFLFCNVNDVCFNFSRNDYSYWLSTPALMPMNPATIGR 60
DB 281 ORAHGODLGTGSCLOPFTMPFLFCNVNDVCFNFSRNDYSYWLSTPALMPMNPATIGR 340

QY 61 ALEPIISRCTV 71
DB 341 ALEPIISRCTV 351

RESULT 4
US-09-167-364-24

Sequence 24, Application US/09167364
Patent No. 6007980
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T
APPLICANT: Morrison, Karen E
APPLICANT: Hudson, Billy G
TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides
FILE REFERENCE: 951263B
CURRENT APPLICATION NUMBER: US/09/167,364
CURRENT FILING DATE: 1998-10-07
EARLIER APPLICATION NUMBER: 08/399889
EARLIER FILING DATE: 1995-03-07
NUMBER OF SEQ ID NOS: 25

SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 24
LENGTH: 471
TYPE: PRT
ORGANISM: Calf
US-09-167-364-24

Query Match
Best Local Similarity 95.1%; Score 371; DB 3; Length 471;
Matches 65; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORAHGODLGTGSCLOPFTMPFLFCNVNDVCFNFSRNDYSYWLSTPALMPMNPATIGR 60
DB 281 ORAHGODLGTGSCLOPFTMPFLFCNVNDVCFNFSRNDYSYWLSTPALMPMNPATIGR 340

QY 61 ALEPIISRCTV 71
DB 341 ALEPIISRCTV 351

RESULT 5
US-08-494-168-2

Sequence 2, Application US/08494168
Patent No. 5731192
GENERAL INFORMATION:
APPLICANT: Reeders, Stephen T.
APPLICANT: Zhou, Jing
TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method
TITLE OF INVENTION: Of Detecting Collagen Deficiency
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/494,168
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/112,465
FILING DATE: 27-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: SAXE, Bernhard D.
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 40397/104/BAAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1694 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-494-168-2

Query Match
Best Local Similarity 36.9%; Score 144; DB 1; Length 1694;
Matches 30; Conservative 10; Mismatches 16; Indels 20; Gaps 2;

QY 14 CLOPFTMPFLFCNVNDVCFNFSRNDYSYWLSTPALMPMNPATIGR 55
DB 1494 CLOPFTMPFLFCNVNDVCFNFSRNDYSYWLSTPALMPMNPATIGR 1551

OY 56 PITGRALEPYISRCTV 71
: : : : :
Db 1552 PVSOTQIPOYISRCV 1567

RESULT 6
US-07-952-853-22
; Sequence 22, Application US/07952853
; Patent No. 5863783
; GENERAL INFORMATION:
; APPLICANT: Van Heuvel, Margaretha
; APPLICANT: Bakhuis, Janna G.
; APPLICANT: Coustel, Yves
; APPLICANT: Harder, Abraham
; APPLICANT: De Graaff, Leendert H.
; APPLICANT: Fliphil, Michel J. A.
; APPLICANT: Van Der Veen, Peter
; APPLICANT: Visser, Jacob
; APPLICANT: Andreoli, Peter M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
; TITLE OF INVENTION: MOLECULES
; TITLE OF INVENTION: ENCODING ARABINAN-DEGRADING ENZYMES OF FUNGAL
; TITLE OF INVENTION: ORIGIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/952,853
; FILING DATE: 19921125
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 246152003500
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-813-5600
; TELEFAX: 415-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 628 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-952-853-22

Query Match 16.2%; Score 63; DB 2; Length 628;
Best Local Similarity 24.6%; Pred. No. 3.6;
Matches 16; Conservative 8; Mismatches 15; Indels 26; Gaps 2;

OY 28 VNDVCFNFSR-----NDYSYWLSTPALMP-----NNAPITGRA 61
: : : : :
Db 263 IGDLCDPRGREGTWYNTDGLGHEFYWCEDLGLVPVGLVWDGFALSGSGNTPLTGDA 322

OY 62 LEPEYI 66
: : : : :
Db 323 LPEYI 327

RESULT 7
US-08-914-848-22
; Sequence 22, Application US/08914848

Patent No. 5989887
; GENERAL INFORMATION:
; APPLICANT: Van Heuvel, Margaretha
; APPLICANT: Bakhuis, Janna G.
; APPLICANT: Coustel, Yves
; APPLICANT: Harder, Abraham
; APPLICANT: De Graaff, Leendert H.
; APPLICANT: Fliphil, Michel J. A.
; APPLICANT: Van Der Veen, Peter
; APPLICANT: Visser, Jacob
; APPLICANT: Andreoli, Peter M.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF DNA
; TITLE OF INVENTION: MOLECULES ENCODING ARABINAN-DEGRADING ENZYMES OF FUNGAL
; TITLE OF INVENTION: ORIGIN
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/914,848
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/952,853
; FILING DATE: 25-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 246152003500
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-813-5600
; TELEFAX: 415-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 628 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-914-848-22

Query Match 16.2%; Score 63; DB 2; Length 628;
Best Local Similarity 24.6%; Pred. No. 3.6;
Matches 16; Conservative 8; Mismatches 15; Indels 26; Gaps 2;

OY 28 VNDVCFNFSR-----NDYSYWLSTPALMP-----NNAPITGRA 61
: : : : :
Db 263 IGDLCDPRGREGTWYNTDGLGHEFYWCEDLGLVPVGLVWDGFALSGSGNTPLTGDA 322

OY 62 LEPEYI 66
: : : : :
Db 323 LPEYI 327

RESULT 8
US-08-766-551-3
; Sequence 3, Application US/08766551
; Patent No. 5840569
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Bandman, Olga
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN GTP-BINDING PROTEINS

NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/766,551
FILING DATE: Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE//DOCKET NUMBER: PP-0168 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 186 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SEQ ID NO: 3
CLONE: 538177
US-08-766-551-3

Query Match 15.4%; Score 60; DB 2; Length 186;
Best Local Similarity 28.8%; Pred. No. 2;
Matches 19; Conservative 8; Mismatches 27; Indels 12; Gaps 2;

OY 5 GODDGLSCGORTTTPFLPCNVNDVCFASRNDYSYL-----STPALPMNM 54
DB 67 GKEL--FSEMDKLMESFVLCVLYDVINESEFNCKWLEKASQAGISLPGVLGNK 124
55 APTGR 60
DB 125 TDLAGR 130

RESULT 9
US-08-098-141-2
Sequence 2, Application US/08098141
Patent No. 5441875
GENERAL INFORMATION:
APPLICANT: Hediger Ph.D., Matthias A. Polypeptide
TITLE OF INVENTION: Urea Transporter Polypeptide
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Choate, Hall & Stewart
STREET: Exchange Place, 53 State Street
CITY: Boston
STATE: MA
COUNTRY: U.S.A.
ZIP: 02119
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/098,141
FILING DATE: 19930723
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kaplan Esq., Warren A.
REGISTRATION NUMBER: 34,199
REFERENCE//DOCKET NUMBER: 092662-001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-227-5020
TELEFAX: 617-227-7566
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 397 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-098-141-2

Query Match 14.6%; Score 57; DB 1; Length 397;
Best Local Similarity 31.4%; Pred. No. 12;
Matches 11; Conservative 6; Mismatches 18; Indels 0; Gaps 0;

OY 34 FASRNDYSYWLSTPALPMNNAPIGTGRALPEYISR 68
DB 135 FSDKDYWWLLLPVIVMSMCPILSSALGTIFSK 169

RESULT 10
US-08-876-991-2
Sequence 2, Application US/08876991
Patent No. 5925360
GENERAL INFORMATION:
APPLICANT: Gregor Meyers, Tillmann R menapf,
APPLICANT: Heinz-J rgen Thiel
TITLE OF INVENTION: Hog cholera virus vaccine and diagnostic
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Organon Teknika Corporation
STREET: 1330-A Piccard Drive
CITY: Rockville
STATE: Maryland
COUNTRY: U.S.A.
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/876,991
FILING DATE: 16-JUN-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/747,577
FILING DATE:
APPLICATION NUMBER: US/08/650,584
FILING DATE:
APPLICATION NUMBER: US/08/469,702
FILING DATE:
APPLICATION NUMBER: US/08/123,596
FILING DATE:
APPLICATION NUMBER: 07/797,554
FILING DATE: 22-NOV-1991
APPLICATION NUMBER: US 07/494,991
FILING DATE: 16-MAR-1990
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: William M. Blackstone
REGISTRATION NUMBER: 29,772
REFERENCE//DOCKET NUMBER:

APPLICANT: Gillespie, Allison
APPLICANT: Feldman, Daniel
APPLICANT: Brenner, Robert
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
METHODS
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Martin, Haller & McClain
STREET: 1660 Union Street
CITY: San Diego
STATE: California
COUNTRY: US
ZIP: 92101
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450,562
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/404,950
FILING DATE: 13-MAR-1995
APPLICATION NUMBER: 08/336,257
FILING DATE: 7-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/314,083
FILING DATE: 28-SEPT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/311,363
FILING DATE: 23-SEPT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/290,012
FILING DATE: 11-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/223,305
FILING DATE: 4-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/193,078
FILING DATE: 07-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/149,097
FILING DATE: 5-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/105,536
FILING DATE: 11-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/914,231
FILING DATE: 13-JULY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/868,354
FILING DATE: 10-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/06903
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/745,206
FILING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/620,250
FILING DATE: 30-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/603,751
FILING DATE: 08-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/482,384
FILING DATE: 02-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US89/01408
FILING DATE: 04-APR-1989
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/176,899
FILING DATE: 04-APR-1988
ATTORNEY/AGENT INFORMATION:
NAME: Seidman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-519812
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 238-0999
TELEFAX: (619) 238-0062
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 520 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-450-562-28

Query Match 14.08; Score 54.5; DB 3; Length 520;
Best Local Similarity 30.28; Pred. No. 36;
Matches 13; Conservative 4; Mismatches 21; Indels 5; Gaps 1;

QY 28 VNDVCFASRNDYSYWLSTPALMPNMNAPITGR-----ALEPY 65
DB 384 LEDACEHLGEYLEAVWRHTTSTPWTPLGRMGSTALSPY 426

RESULT 14
US-08-976-255-17
Sequence 17, Application US/08976255
Patent No. 6136581
GENERAL INFORMATION:
APPLICANT: Jono, Keith E.
APPLICANT: Plozman, Gregory
TITLE OF INVENTION: KINASE GENES AND USES
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FASTSEQ for Windows 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976,255
FILING DATE: No. 6136581ember 21, 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/031,675
FILING DATE: No. 6136581ember 22, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 229/182
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-976-255-17

THIS PAGE BLANK (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2001, 15:41:17 ; Search time 49.95 Seconds

(without alignments)
81.253 Million cell updates/sec

Title: US-09-543-371-10_COPY_54_124

Perfect score: 390

Sequence: 1 QRAHGQDGLTGLSCLOQRFMT.....MNMPIITGRALEPIYSRIV 71

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_0401.*
1: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	390	100.0	211	21	Y95918 Human Goodpasture Partial sequence o
2	390	100.0	218	16	R79164 Human type IV coll
3	390	100.0	218	20	Y44172 Human alpha3 type
4	390	100.0	218	21	Y56784 Human type IV coll
5	390	100.0	245	21	Y67942 Human type IV coll
6	390	100.0	268	20	Y31993 Human alpha3(IV)NC
7	390	100.0	268	21	Y97555 Human alpha3(IV)NC
8	371	95.1	471	20	Y44171 Bovine type IV col
9	371	95.1	471	21	Y56783 Bovine alpha3 type
10	369	94.6	471	16	R79163 Partial sequence o
11	321	82.3	229	21	Y67943 Human type IV coll

12	321	82.3	260	20	Y31991 Type IV collagen N
13	321	82.3	260	21	Y97553 Human alpha1(IV)NC
14	321	82.3	406	21	B58169 Lung cancer associ
15	315	80.8	229	10	P93524 Complete sequence
16	314	80.5	264	20	Y31995 Type IV collagen N
17	314	80.5	264	21	Y97557 Human alpha5(IV)NC
18	314	80.5	309	21	B54044 Human pancreatic c
19	314	80.5	772	13	R23873 Human alpha 5 (IV)
20	314	80.5	772	18	W09643 Human type IV coll
21	242	62.1	242	21	Y67946 Human type IV coll
22	242	62.1	258	20	Y31992 Type IV collagen N
23	242	62.1	258	21	Y97554 Human alpha2(IV)NC
24	242	62.1	260	20	B58180 Lung cancer associ
25	235	60.3	260	20	Y31996 Type IV collagen N
26	235	60.3	260	21	Y97558 Human alpha6(IV)NC
27	222	56.9	260	20	Y31994 Type IV collagen N
28	222	56.9	260	21	Y97556 Human alpha4(IV)NC
29	144	36.9	1694	19	W40109 Human alpha-6(IV)
30	63	16.2	628	13	R27575 ABF-A from A. nige
31	62.5	16.0	252	17	R95251 DNA mismatch repa
32	60	15.4	186	19	W57884 Human GTP-binding
33	60	15.4	1495	18	W18226 Transcriptional co
34	60	15.4	2518	21	B40574 Human ORFX ORF38
35	59	15.1	160	13	R27887 Pheromone receptor
36	58	14.9	1445	20	W93595 O. longistaminata
37	57.5	14.7	397	16	W13742 Rat pheromone rece
38	57	14.6	256	17	R95250 Urea transporter p
39	56.5	14.5	256	17	R95250 DNA mismatch repa
40	56.5	14.5	5532	21	B23752 S. avermilitis ave
41	55.5	14.2	174	18	W11722 H-Delta-1 polypept
42	55	14.1	271	21	G28508 Arabidopsis thalia
43	55	14.1	271	21	G43705 Arabidopsis thalia
44	55	14.1	300	21	G28507 Arabidopsis thalia
45	55	14.1	300	21	G43704 Arabidopsis thalia

ALIGNMENTS

RESULT 1	
Y95918	standard; Protein; 211 AA.
XX	
AC	Y95918:
XX	
DT	20-NOV-2000 (first entry)
XX	
DE	Human Goodpasture antigen Deltav.
XX	
KW	Goodpasture antigen; GPdeltav; goodpasture antigen binding protein;
KW	GPBP; human; autoimmune disease; apoptosis; cancer; tumour;
KW	therapy.
XX	
OS	Homo sapiens.
XX	
PN	W0200050607-A2.
XX	
PD	31-AUG-2000.
XX	
PF	24-FEB-2000; 2000WO-1B00324.
XX	
PR	24-FEB-1999; 99US-0121483.
XX	
PA	(SAUS/) SAUS J.
XX	
PI	Saus J;
XX	
DR	WPI: 2000-572094/53.
XX	N-PSDB; A50367.
PT	Novel Goodpasture antigen binding proteins useful for diagnosing and
XX	treating autoimmune disorders, tumor, and preventing cell apoptosis

PS Claim 36: Page 151-152; 158pp; English.

XX The present sequence is that of human recombinant Goodpasture
 CC antigen (GP) Deltav, i.e. an alternative form of human GP resulting
 CC from splicing out of exon V. The recombinant protein, lacking the
 CC Met-1 residue, was expressed in bacterial pellets using modified
 CC vector pET15b carrying GpDeltav cDNA (see A50367). The invention
 CC relates to novel Goodpasture antigen binding proteins (GPRPs; see
 CC Y93900-11), which bind to and phosphorylate the unique N-terminal
 CC region of human GP, and which are highly expressed in several
 CC autoimmune conditions. Claimed methods for treating an autoimmune
 CC disorder, cell apoptosis or a tumour involve modifying the
 CC expression or activity of GPRP, especially using a GP-derived
 CC peptide, such as GpDeltav.

SQ Sequence 211 AA:

Query Match 100.0%; Score 390; DB 21; Length 211;
 Best Local Similarity 100.0%; Pred. No. 2,6e-43;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORAHGODLGTGSCLOFTTTPFLFCNVNDVCNPFASRNDYSYWLSTPALPMNMAPITGR 60
 DB 54 qrahgqdlgtlgsclqftrftmpflfcnvndvcnfasrndysywlstpalpmnmapitgr 113

QY 61 ALEPIYSRCTV 71
 DB 114 alepiysrctv 124

RESULT 2

ID R79164 standard; Protein: 218 AA.

AC R79164;
 XX
 XX 22-DEC-1995 (first entry)
 DE Partial sequence of human alpha 3 chain of type IV collagen.
 XX
 XX Type IV collagen; alpha 3 chain; Alport syndrome; COL4A3 gene.
 KM
 XX Homo sapiens.
 OS
 XX
 XX USS424408-A.
 PN
 XX 13-JUN-1995.
 PD
 XX 30-NOV-1990; 90US-0621091.
 PR
 XX 30-NOV-1990; 90US-0621091.
 PA (UNITV) UNIV KANSAS MEDICAL CENT.
 XX (UYVA) UNIV YALE.
 XX
 XX Hudson BG, Morrison KE, Reeders ST;
 PI
 XX WPI: 1995-262631/34.
 DR N-PSDB; Q96291.
 DR
 XX
 XX cDNAs encoding human or bovine alpha-3 type 4 collagen peptide(s)
 PT useful for detection and therapeutic removal of antibodies associated
 PT with Goodpasture syndrome
 PS Disclosure; Columns 7-10; 33pp; English.

XX Using the PCR with primers derived from each end of the known 27 AA
 CC residue bovine alpha 3 (IV) collagen protein sequence, a 68 bp
 CC bovine genomic fragment was amplified. This fragment was then used
 CC to a bovine lens cDNA library and a 1.5 kb partial cDNA clone was
 CC obt'd. (clone KMC15). This encodes 238 residues of the triple helical
 CC collagenous domain and all 233 residues of the C-terminal non-

CC collagenous (NC1) domain of the alpha 3 (IV) chain. This bovine
 CC cDNA clone was used to screen a human kidney cDNA library and a
 CC 2.7 kb human cDNA clone (clone KMC27) was obt'd. This clone encodes
 CC 218 residues of the NC1 domain and a portion of the 3' UTR region
 CC of the human alpha 3(IV) chain. The COL4A3 gene localises to
 CC chromosome 2 and therefore mutations in COL4A3 cannot be
 CC responsible for Alport syndrome which is X-linked. An isolated
 CC and substantially pure nt. having the sequence in Q96291
 CC is claimed.

SQ Sequence 218 AA:

Query Match 100.0%; Score 390; DB 16; Length 218;
 Best Local Similarity 100.0%; Pred. No. 2,7e-43;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORAHGODLGTGSCLOFTTTPFLFCNVNDVCNPFASRNDYSYWLSTPALPMNMAPITGR 60
 DB 28 qrahgqdlgtlgsclqftrftmpflfcnvndvcnfasrndysywlstpalpmnmapitgr 87

QY 61 ALEPIYSRCTV 71
 DB 88 alepiysrctv 98

RESULT 3

ID Y44172 standard; Protein: 218 AA.

AC Y44172;
 XX
 XX 01-FEB-2000 (first entry)
 DE Human type IV collagen alpha3 chain protein.
 XX
 XX Recombinant; bovine; alpha3 chain; type IV collagen; detection;
 KM Goodpasture syndrome; antibody; blood; tissue; human; nephrotrophism.
 XX
 XX Homo sapiens.
 OS
 XX
 XX USS973120-A.
 PN
 XX 26-OCT-1999.
 PD
 XX 07-MAR-1995; 95US-0399889.
 PR
 XX 30-NOV-1990; 90US-0621091.
 PA (UYVA) UNIV YALE.
 XX (UNITV) UNIV KANSAS MEDICAL CENT.
 XX
 XX Hudson BG, Reeders ST, Morrison KE;
 PI
 XX WPI: 1999-610317/52.
 DR N-PSDB; Z28775.
 DR
 XX
 XX Isolated alpha 3 chain of type IV collagen polypeptide useful for
 PT diagnosis and treatment of Goodpasture syndrome -
 PT
 PS Claim 2; Column 35-36; 27pp; English.

XX This sequence represents a recombinant human alpha3 chain of type IV
 CC collagen polypeptide. The sequence corresponds to the 218 amino acids
 CC of the C-terminal non-collagenous domain. Alpha3 chain collagen
 CC polypeptides are useful for detecting Goodpasture antibodies in blood
 CC or tissue from a human patient and for treating Goodpasture syndrome,
 CC especially by neutralising the antibodies in the blood. The polypeptides
 CC also have a nephrotrophic activity.

SQ Sequence 218 AA:

Query: Match 100.0%; Score 390; DB 20; Length 218;
Best Local Similarity 100.0%; Pred. No. 2.7e-43;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORAHGDDLTGSCLOPFTTPEFLFCNVNDVCFASRNDYSYWLSPALPMNMAPITGR 60
DB 28 grahgddlgltgscldqrlftmpflfcnvndvcnfasrndysywlspalpmnmapitgr 87

OY 61 ALEPYISRCTV 71
DB 88 alepyisrctv 98

RESULT 4
Y56784
ID Y56784 standard; Protein: 218 AA.

AC Y56784;
XX
XX 27-MAR-2000 (first entry)

Human alpha3 type IV collagen C-terminal domain.

XX
XX Goodpasture syndrome; type IV collagen; alpha3 chain; human.
XX
XX Homo sapiens.
XX
XX US6007980-A.
XX
XX 28-DEC-1999.
XX
XX 07-OCT-1998; 98US-0167364.
XX
XX 30-NOV-1990; 90US-0621091.
XX
XX 07-MAR-1995; 95US-0399889.
XX
XX (UNIV) UNIV KANSAS MEDICAL CENT.
XX
XX (UYVA) UNIV YALE.

PI Hudson BG, Reeders ST, Morrison KE;
XX
XX WPI: 2000-096371/08.
XX
XX N-PSDB: 246729.

PT Diagnosing and treating Goodpasture syndrome using a peptide derived
PT from type IV collagen
XX
XX
XX Disclosure: Columns 23-26; 26pp; English.

CC The invention provides a method of detecting Goodpasture antibodies
CC in the fluid of a patient by contacting it with a peptide comprising at
CC most 218 amino acids of the human alpha3 chain type IV collagen that
CC contains the fragment shown in Y56785. The methods are useful for the
CC diagnosis and treatment of Goodpasture syndrome. The present sequence
CC represents the carboxy terminal noncollagenous domain of the human
CC alpha3 chain of type IV collagen.
XX
XX
XX Sequence 218 AA;

Query Match 100.0%; Score 390; DB 21; Length 218;
Best Local Similarity 100.0%; Pred. No. 2.7e-43;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORAHGDDLTGSCLOPFTTPEFLFCNVNDVCFASRNDYSYWLSPALPMNMAPITGR 60
DB 28 grahgddlgltgscldqrlftmpflfcnvndvcnfasrndysywlspalpmnmapitgr 87

OY 61 ALEPYISRCTV 71
DB 88 alepyisrctv 98

RESULT 5
Y67942
ID Y67942 standard; Protein: 245 AA.

AC Y67942;
XX
XX 03-APR-2000 (first entry)

Human type IV collagen alpha 3 chain protein sequence SEQ ID NO:10.

XX
XX Human; type IV collagen; anti-angiogenic; angiogenesis; cancer;
XX benign tumour; rheumatoid arthritis; diabetic retinopathy; psoriasis;
XX ocular angiogenesis disease; Osler-Weber Syndrome; telangiectasia;
XX myocardial angiogenesis; plaque neovascularisation; angiodioma;
XX atherosclerosis; scleroderma; hypertrophic scar; cat scratch disease;
XX contraception; obesity.

XX
XX Homo sapiens.
XX
XX WO9965940-A1.
XX
XX 23-DEC-1999.
XX
XX 17-JUN-1999; 99WO-US13737.
XX
XX 17-JUN-1998; 98US-0089689.
XX
XX 25-MAR-1999; 99US-0126175.
XX
XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
XX Kalluri R;
XX
XX WPI: 2000-097708/08.
XX
XX N-PSDB: Z57158.

DR Anti-angiogenic proteins comprising the NC1 domain of the alpha 1, 2 or
DR 3 chain of type IV collagen used in, e.g. treatment of benign tumors
XX and rheumatoid arthritis
XX
XX
XX Claim 32; Fig 16B; 117pp; English.

PS The present sequence represents the human type IV collagen alpha 3 chain.
XX
XX The present invention describes an isolated protein chosen from the NC1
XX domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or
XX a fragment, analogue, derivative or mutant, which has anti-angiogenic
XX properties. The anti-angiogenic proteins, multimers and chimeras are
XX useful for inhibiting angiogenic activity in mammalian tissue.
XX
XX especially for treating diseases chosen from angiogenesis-dependent
XX cancers, benign tumours, rheumatoid arthritis, diabetic retinopathy,
XX psoriasis, ocular angiogenesis diseases, Osler-Weber Syndrome,
XX myocardial angiogenesis, plaque neovascularisation, telangiectasia,
XX haemophilic joints, angiodioma, wound granulation, intestinal
XX adhesions, atherosclerosis, scleroderma, hypertrophic scars, cat scratch
XX disease, Helicobacter pylori ulcers, dialysis graft vascular access
XX stenosis, contraception and obesity. The compositions can be used to
XX inhibit a disease characterised by angiogenic activity. In conjunction
XX with radiation therapy, chemotherapy or immunotherapy.

XX
XX Sequence 245 AA;

Query Match 100.0%; Score 390; DB 21; Length 245;
Best Local Similarity 100.0%; Pred. No. 3.1e-43;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORAHGDDLTGSCLOPFTTPEFLFCNVNDVCFASRNDYSYWLSPALPMNMAPITGR 60
DB 55 grahgddlgltgscldqrlftmpflfcnvndvcnfasrndysywlspalpmnmapitgr 114

OY 61 ALEPYISRCTV 71
DB 115 alepyisrctv 125

RESULT	6
XX	Y31993
XX	Y31993 standard; Protein; 268 AA.
XX	Y31993;
XX	05-JAN-2000 (first entry)
DE	Type IV collagen NCI domain alpha-3 monomer.
XX	
KW	Type IV collagen; NCI domain; non-collagenous domain; human;
KW	angiogenesis; tumour; metastasis; therapy; diabetic retinopathy;
KW	rheumatoid arthritis; retinal neovascularization;
KW	choroidal neovascularization; macular degeneration;
KW	corneal neovascularization; retinopathy of prematurity;
KW	corneal graft rejection; neovascular glaucoma;
KW	retrolental fibroplasia; epidemic keratoconjunctivitis;
KW	vitamin A deficiency; contact lens overwear; atopic keratitis;
KW	superior limbic keratitis; pterygium keratitis sicca; sogrens;
KW	acne rosacea; phlyctenulosis; syphilis; Mycobacteria infection;
KW	lipid degeneration; chemical burn; ulcer; herpes simplex infection;
KW	Herpes zoster infection; protozoan infection; kaposi's sarcoma;
KW	Mooren ulcer; Terrien's marginal degeneration;
KW	marginal keratolysis; trachoma; systemic lupus; polyarteritis;
KW	Wegener's sarcoidosis; scleritis; Steven's Johnson disease;
KW	radial keratotomy; sickle cell anaemia; sarcoid;
KW	pseudophtholoma elasticum; Paget's disease; vein occlusion;
KW	artery occlusion; carotid obstructive disease; chronic uveitis;
KW	chronic vitritis; Lyme's disease; Eales disease; bechets disease;
KW	myopia; optic pty; Stargart's disease; pars planitis;
KW	chronic retinal detachment; hyperviscosity syndrome; toxoplasmosis;
KW	post-laser complication; fibrovascular tissue proliferation;
KW	haemangioma; Osler-Weber-Rendu; AIDS; ocular neovascular disease;
KW	osteoarthritis; chronic inflammation; Crohn's disease;
KW	ulcerative colitis; psoriasis; atherosclerosis; pemphigoid.
OS	Homo sapiens.
OS	Synthetic.
XX	
FH	Location/Qualifiers
FT	Key
FT	Peptide
FT	Protein
FT	Peptide
FT	Protein
FT	Peptide
FT	Protein
XX	
PN	WO9449885-A2.
XX	
PD	07-OCT-1999.
XX	
PF	26-MAR-1999; 99WO-US06445.
XX	
PR	27-MAR-1998; 98US-0079783.
XX	
PR	29-OCT-1998; 98US-0106170.
XX	
PA	(UNIV) UNIV KANSAS MEDICAL CENT.
PI	Hudson BG, Sarraa MP;
XX	
DR	WPI: 1999-601297/51.
XX	
DR	N-PSDB; Z20091.
XX	
PT	Inhibition of angiogenesis with non-collagenous alpha chain monomer
XX	useful for treating e.g. tumor growth or metastasis,
XX	neovascularisation, etc -
XX	
PS	Disclosure; Fig 17c; 56pp; English.
CC	This sequence represents a recombinant type IV collagen

CC	non-collagenous (NC1) domain alpha-3 polypeptide composed of a
CC	BM40 signal sequence (which is cleaved from the mature protein) to
CC	facilitate protein secretion, and a mature protein comprising an
CC	affinity tag (facilitates purification and identification of the
CC	material) and the alpha-1 chain monomer. The invention provides
CC	methods and kits for inhibiting angiogenesis, tumour growth and
CC	metastasis, and endothelial cell interaction with the extracellular
CC	matrix, each method comprising contacting the tumour or animal
CC	tissue with 1 or more isolated type IV collagen NC1 alpha chain
CC	monomer(s) selected from the group consisting of alpha-1, alpha-2,
CC	alpha-3 and alpha-6 NC1 chain monomers (see 131991-96). The
CC	monomers can be produced via recombinant protein expression. The
CC	polynucleotides and polypeptides are used to treat an angiogenesis-
CC	mediated disorder or condition, especially selected from solid and
CC	blood-borne tumours, diabetic retinopathy, rheumatoid arthritis,
CC	retinal neovascularization, choroidal neovascularization, macular
CC	degeneration, corneal neovascularization, retinopathy of prematurity,
CC	corneal graft rejection, neovascular glaucoma, retrolental
CC	fibroplasia, epidemic keratoconjunctivitis, vitamin A deficiency,
CC	contact lens overwear, atopic keratitis, superior limbic keratitis,
CC	pterygium keratitis sicca, seborrheic acne rosacea, phlyctenulosis,
CC	syphilis, mycobacteria infections, lipid degeneration, chemical
CC	burns, bacterial ulcers, fungal ulcers, herpes simplex infections,
CC	herpes zoster infections, protozoan infections, kaposi's sarcoma,
CC	Mooren ulcer, kerien's marginal degeneration, marginal keratolysis,
CC	trauma, systemic lupus, polyarteritis, Wegener's sarcoidosis,
CC	scleritis, Steven's Johnson disease, radial keratotomy, sickle cell
CC	anemia, sarcoid, pseudoxanthoma elasticum, Pagets disease, vein
CC	occlusion, artery occlusion, carotid obstructive disease, chronic
CC	uveitis, chronic vitritis, Lyme's disease, Fales disease, Behcets
CC	disease, myopia, optic pits, Stargatts disease, pars planitis,
CC	chronic retinal detachment, hyperviscosity syndromes, toxoplasmosis,
CC	post-laser complications, abnormal proliferation of fibrovascular
CC	tissue, haemangiomas, Osler-Weber-Rendu, AIDS, ocular neovascular
CC	disease, osteoarthritis, chronic inflammation, Crohn's disease,
CC	ulcerative colitis, psoriasis, atherosclerosis, and pemphigoid (all
CC	claimed).
XX	
XX	Sequence 268 AA;
QQ	
Query Match	100.0%; Score 390; DB 20; Length 268;
Best Local Similarity	100.0%; Pred. No. 3 4e-43;
Matches 71: Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 QRAHQDGLTGSCIQRTTTPLEFCNVNDVCFNPSARNDYSYWLSTPALMPNNAPTGR 60
DB	
78	granggdglglgscilgrfttppflfcnvndvcnfasrindgywistpalmnnaptlgr 137
DB	61 ALEPISRCTV 71
138	alepissrcv 148
RESULT	7
ID	Y97555
AC	Y97555 standard; protein; 268 AA.
XX	Y97555;
XX	
DT	12-FEB-2001 (first entry)
XX	
DE	Human alpha3(IV)NC1 protein sequence.
XX	
XX	Type IV collagen alpha chain monomer; human; inhibitor; angiogenesis;
KW	tumour growth; integrin receptor; carcinoma; sarcoma; rhadomyosarcoma;
KW	retinoblastoma; Ewing sarcoma; neuroblastoma; osteosarcoma; leukaemia;
KW	diabetic retinopathy; rheumatoid arthritis; neovascularisation;
KW	muscular degeneration; corneal graft rejection; vitamin A deficiency;
KW	atopic keratitis; mycobacteria infection; chemical burn; sarcoid;
KW	kaposi's sarcoma; sickle cell anemia; carotid obstructive disease;
KW	chronic inflammation; psoriasis; therapy; alpha3(IV)NC1.

OS Homo sapiens.
 XX
 PN W0200059532-A1.
 XX
 PD 12-OCT-2000.
 XX
 PF 31-MAR-2000; 2000WO-US08678.
 XX
 PR 01-APR-1999; 99US-0127391.
 XX
 PA (BIOS-) BIOSTRATUM INC.
 XX
 PI Brooks P, Hudson B;
 XX
 DR WPI: 2000-664962/64.
 DR N-PSDB: A90993.
 XX
 PT Use of antagonists of specific integrin receptors for inhibiting
 PT angiogenesis, tumour growth or metastases, or endothelial cell
 PT interactions with the extracellular matrix -
 XX
 P1 Disclosure; Fig 17c; 78bp; English.
 XX
 CC This sequence is a human type IV collagen alpha chain monomer,
 CC designated alpha3(IV)N1. The invention relates to a method for
 CC inhibiting angiogenesis, tumour growth or metastases, or endothelial cell
 CC interactions with the extracellular matrix, comprising contacting the
 CC cells or tissue with a polypeptide composition containing antagonists of
 CC specific integrin receptors. The methods and the antagonists are useful
 CC for inhibiting angiogenesis, tumour growth or metastases, or endothelial
 CC cell interaction with the extracellular matrix. The antagonists are also
 CC useful for treating diseases and conditions with accompanying undesired
 CC angiogenesis, e.g. solid and blood-borne tumours (e.g. melanomas,
 CC carcinomas, sarcomas, rhabdomyosarcoma, retinoblastoma, Ewing sarcoma,
 CC neuroblastoma, osteosarcoma or leukaemia). These are also applicable to
 CC treating non-tumorigenic diseases and conditions with accompanying
 CC undesired angiogenesis, e.g. diabetic retinopathy, rheumatoid arthritis,
 CC retinal neovascularisation, choroidal neovascularisation, muscular
 CC degeneration, corneal graft rejection, vitamin A deficiency, atopic
 CC keratitis, Mycobacteria infections, chemical burns, Kaposi's sarcoma,
 CC sickle cell anaemia, sarcoid, carotid obstructive disease, post-laser
 CC complications, chronic inflammation or psoriasis.
 CC
 SO Sequence 268 AA;

Query Match 100.0%; Score 390; DB 21; Length 268;
 Best Local Similarity 100.0%; Pred. No. 3.4e-43;
 Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORAHODGTGSCIORFTMPLEFCNVNDVCFNFSRDYSYWLSTPALMPNNAPITGR 60
 DB 78 qrahgqdgltgsciqrlttmpfllcnvndvcnfasrindysywlstpalmpnnapitgr 137
 QY 61 ALEPIRSCTV 71
 DB 138 alepysrctv 148

RESULT 8
 Y44171
 ID Y44171 standard; Protein; 471 AA.
 XX
 AC Y44171;
 XX
 DT 01-FEB-2000 (first entry)
 XX
 DE Bovine type IV collagen alpha3 chain protein.
 XX
 KW Recombinant; bovine; alpha3 chain; type IV collagen; detection;
 KM Goodpasture syndrome; antibody; blood; tissue; human; nephrotrophism.
 XX
 OS Bos taurus.

XX US5973120-A.
 PN
 XX 26-OCT-1999.
 PD
 XX 07-MAR-1995; 95US-0399889.
 PF
 XX 30-NOV-1990; 90US-0621091.
 PR
 XX (UYVA) UNIV YALE.
 PA (UNIV) UNIV KANSAS MEDICAL CENT.
 XX
 PI Hudson BG, Reeders ST, Morrison KE;
 XX
 DR WPI: 1999-610317/52.
 DR N-PSDB: Z28774.
 XX
 PT Isolated alpha 3 chain of type IV collagen polypeptide useful for
 PT diagnosis and treatment of Goodpasture syndrome -
 XX
 PS Claim 1; Column 31-34; 27pp; English.
 XX
 CC This sequence represents a recombinant bovine alpha3 chain of type IV
 CC collagen polypeptide. The sequence corresponds to the 238 amino acids
 CC of the C-terminal end of the triple helical domain and all 233 amino
 CC acids of the C-terminal non-collagenous domain. Alpha3 chain collagen
 CC polypeptides are useful for detecting Goodpasture antibodies in blood
 CC or tissue from a human patient and for treating Goodpasture syndrome,
 CC especially by neutralising the antibodies in the blood. The polypeptides
 CC also have a nephrotrophic activity.
 CC
 SO Sequence 471 AA;

Query Match 95.1%; Score 371; DB 20; Length 471;
 Best Local Similarity 91.5%; Pred. No. 2.1e-40;
 Matches 65; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORAHODGTGSCIORFTMPLEFCNVNDVCFNFSRDYSYWLSTPALMPNNAPITGR 60
 DB 261 eqahgqdgltgsciqrlttmpfllcnvndvcnfasrindysywlstpalmpnnapitgr 340
 QY 61 ALEPIRSCTV 71
 DB 341 alepysrctv 351

RESULT 9
 Y56783
 ID Y56783 standard; Protein; 471 AA.
 XX
 AC Y56783;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Bovine alpha3 type IV collagen.
 XX
 KW Goodpasture syndrome; type IV collagen; alpha3 chain; bovine.
 KM
 XX
 OS Bos sp.
 XX
 PN US6007980-A.
 XX
 PD 28-DEC-1999.
 XX
 PF 07-OCT-1998; 98US-0167364.
 XX
 PR 30-NOV-1990; 90US-0621091.
 PR 07-MAR-1995; 95US-0399889.
 XX
 PA (UNIV) UNIV KANSAS MEDICAL CENT.
 PA (UYVA) UNIV YALE.
 XX

```

PI Hudson BG, Reeders ST, Morrison KE;
XX WPI; 2000-096371/08.
DR N-PSDB; 246728.
XX
PT Diagnosing and treating Goodpasture syndrome using a peptide derived
PT from type IV collagen
XX
PS Disclosure; Columns 19-24; 26pp; English.
XX
CC The invention provides a method of detecting Goodpasture antibodies
CC in the fluid of a patient by contacting it with a peptide comprising at
CC most 218 amino acids of the human alpha3 chain type IV collagen that
CC contains the fragment shown in Y56785. The methods are useful for the
CC diagnosis and treatment of Goodpasture syndrome. The present sequence
CC represents the bovine alpha3 chain of type IV collagen.
XX
SQ Sequence 471 AA;

Query Match
Best Local Similarity 95.1%; Score 371; DB 21; Length 471;
Matches 65; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRAHGQDLGTLGSLQRFRTTLPFLFCNVNDVNCNFSANDSYSLSTPALMPMNAPIITGR 60
   :|||||
Db 281 egahgqdlgtlgsclgrftftmpflfcncfnasrndsywslstpalmpmdnapltgr 340
   :|||||

QY 61 ALEPYISRCTV 71
   :|||||
Db 341 alepyisrcv 351

RESULT 10
ID R79163
AC R79163 standard; Protein; 471 AA.
XX
XX R79163;
XX
DT 22-DEC-1995 (first entry)
XX
DE Partial sequence of bovine alpha 3 chain of type IV collagen.
XX
XX Type IV collagen; alpha 3 chain.
XX
XX Bos taurus.
XX
XX US5424408-A.
XX
XX 13-JUN-1995.
XX
XX 30-NOV-1990; 90US-0621091.
XX
XX 30-NOV-1990; 90US-0621091.
XX
XX 30-NOV-1990; 90US-0621091.
XX
XX (UNIV ) UNIV KANSAS MEDICAL CENT.
XX
XX (UYVA ) UNIV YALE.
XX
XX Hudson BG, Morrison KE, Reeders ST;
XX
XX WPI; 1995-262631/34.
XX
XX N-PSDB; Q96290.
XX
XX CDNA encoding human or bovine alpha-3 type 4 collagen peptide(s)
XX useful for detection and therapeutic removal of antibodies associated
XX with Goodpasture syndrome
XX
XX Disclosure; Columns 5-8; 33pp; English.
XX
XX using the PCR with primers derived from each end of the known 27 AA
XX residue bovine alpha 3 (IV) collagen protein sequence, a 68 bp
XX bovine genomic fragment was amplified. This fragment was then used
XX to a bovine lens cDNA library and a 1.5 kb partial cDNA clone was

```

```

CC obtd. (clone KMC15). This encodes 238 residues of the triple helical
CC collagenous domain and all 233 residues of the C-terminal non-
CC collagenous (NC1) domain of the alpha 3 (IV) chain. An isolated and
CC substantially pure nt. having the sequence in Q96290 is claimed.
XX
SQ Sequence 471 AA;

Query Match
Best Local Similarity 94.6%; Score 369; DB 16; Length 471;
Matches 64; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRAHGQDLGTLGSLQRFRTTLPFLFCNVNDVNCNFSANDSYSLSTPALMPMNAPIITGR 60
   :|||||
Db 281 egahgqdlgtlgsclgrftftmpflfcncfnasrndsywslstpalmpmdnapltgr 340
   :|||||

QY 61 ALEPYISRCTV 71
   :|||||
Db 341 alepyisrcv 351

RESULT 11
ID Y67943
AC Y67943 standard; Protein; 229 AA.
XX
XX Y67943;
XX
DT 03-APR-2000 (first entry)
XX
DE Human type IV collagen alpha 1 chain protein sequence SEQ ID NO:2.
XX
XX Human; type IV collagen; anti-angiogenic; angiogenesis; cancer;
XX benign tumour; rheumatoid arthritis; diabetic retinopathy; psoriasis;
XX ocular angiogenesis disease; Osler-Weber Syndrome; telangiectasia;
XX myocardial angiogenesis; plaque neovascularisation; angiodioma;
XX atherosclerosis; scleroderma; hypertrophic scar; cat scratch disease;
XX contraception; obesity.
XX
XX Homo sapiens.
XX
XX OS
XX PN WO9965940-A1.
XX
XX PD 23-DEC-1999.
XX
XX PF 17-JUN-1999; 99WO-US13737.
XX
XX PR 17-JUN-1998; 98US-0089689.
XX
XX PR 25-MAR-1999; 99US-0126175.
XX
XX PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
XX
XX PI Kalluri R;
XX
XX WPI; 2000-097708/08.
XX
XX DR N-PSDB; 257159.
XX
XX PT Anti-angiogenic proteins comprising the NC1 domain of the alpha 1, 2 or
XX 3 chain of type IV collagen used in, e.g. treatment of benign tumors
XX and rheumatoid arthritis -
XX
XX Example 1; Fig 1B; 117pp; English.
XX
XX The present sequence represents the human type IV collagen alpha 1 chain.
XX The present invention describes an isolated protein chosen from the NC1
XX domain of the alpha 1, alpha 2 or alpha 3 chains of type IV collagen or
XX a fragment, analogue, derivative or mutant, which has anti-angiogenic
XX properties. The anti-angiogenic proteins, multimers and chimeras are
XX useful for inhibiting angiogenic activity in mammalian tissue,
XX especially for treating diseases chosen from angiogenesis-dependent
XX cancers, benign tumours, rheumatoid arthritis, diabetic retinopathy,
XX psoriasis, ocular angiogenesis diseases, Osler-Weber Syndrome,
XX myocardial angiogenesis, plaque neovascularisation, telangiectasia,
XX haemophilic joints, angiodioma, wound granulation, intestinal

```


RESULT 13
ID Y97553 standard; Protein: 260 AA.
XX Y97553;
AC XX
DT 12-FEB-2001 (first entry)
XX XX
DE Human alpha1(IV)NC1 protein sequence.
XX XX
KW Type IV collagen alpha chain monomer; human; inhibitor; angiogenesis;
KW tumour growth; integrin receptor; carcinoma; sarcoma; rhabdomyosarcoma;
KW retinoblastoma; Ewing sarcoma; neuroblastoma; osteosarcoma; leukemia;
KW diabetic retinopathy; rheumatoid arthritis; neovascularisation;
KW muscular degeneration; corneal graft rejection; vitamin A deficiency;
KW atopic keratitis; Mycobacteria infection; chemical burn; sarcoid;
KW Kaposi's sarcoma; sickle cell anaemia; carotid obstructive disease;
KW chronic inflammation; psoriasis; therapy; alpha1(IV)NC1.
XX XX
OS Homo sapiens.
XX XX
PN WO200059532-A1.
XX XX
PD 12-OCT-2000.
XX XX
PF 31-MAR-2000; 2000WO-US08678.
XX XX
PR 01-APR-1999; 99US-0127391.
XX XX
PA (BIO-) BIOSTRATUM INC.
XX XX
PI Brooks P, Hudson B;
XX XX
DR WPI: 2000-664962/64.
XX XX
DR N-PSDB; A90991.
XX XX
PT Use of antagonists of specific integrin receptors for inhibiting
PT angiogenesis, tumour growth or metastases, or endothelial cell
PT interactions with the extracellular matrix -
XX XX
PS Disclosure; Fig 17a; 78pp; English.
XX XX
CC This sequence is a human type IV collagen alpha chain monomer,
CC designated alpha1(IV)NC1. The invention relates to a method for
CC inhibiting angiogenesis, tumour growth or metastases, or endothelial cell
CC interactions with the extracellular matrix, comprising contacting the
CC cells or tissue with a polypeptide composition containing antagonists of
CC specific integrin receptors. The methods and the antagonists are useful
CC for inhibiting angiogenesis, tumour growth or metastases, or endothelial
CC cell interaction with the extracellular matrix. The antagonists are also
CC useful for treating diseases and conditions with accompanying undesired
CC angiogenesis, e.g. solid and blood-borne tumours (e.g. melanomas,
CC carcinomas, sarcomas, rhabdomyosarcoma, retinoblastoma, Ewing sarcoma,
CC neuroblastoma, osteosarcoma or leukaemia). These are also applicable to
CC treating non-tumorigenic diseases and conditions with accompanying
CC undesired angiogenesis, e.g. diabetic retinopathy, rheumatoid arthritis,
CC retinal neovascularisation, choroidal neovascularisation, muscular
CC degeneration, corneal graft rejection, vitamin A deficiency, atopic
CC keratitis, Mycobacteria infections, chemical burns, Kaposi's sarcoma,
CC sickle cell anaemia, sarcoid, carotid obstructive disease, post-laser
CC complications, chronic inflammation or psoriasis.
XX XX
SQ Sequence 260 AA.

Query Match 82.3%; Score 321; DB 21; Length 260;
Best Local Similarity 77.5%; Pred. No. 3,7e-34;
Matches 55; Conservative 9; Mismatches 7; Indels 0; Gaps 0;
QY 1 ORAHGDICTGSGCLQRFITMPFLFCNVNDVCNFAASRNDYSYWLSPALPMNMNAPTGR 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

Db 71 erahgdiqltag sclrkfstmpflfcnlunvncfnasrndysywlstpepnmmapltge 130
QY 61 ALEPYISRCTV 71
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 131 nirflisrcav 141
RESULT 14
ID B58169 standard; Protein: 406 AA.
XX B58169;
AC XX
DT 14-MAR-2001 (first entry)
XX XX
DE Lung cancer associated polypeptide sequence SEQ ID 507.
XX XX
KW Human; lung cancer associated protein; neuroprotective; cytoskeletal;
KW cardioactive; immunomodulatory; muscular active; vulnerrary;
KW gastrointestinal; nephrotoxic; antiinfective; gynecological;
KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
KW proliferative disorder; wound healing; infectious disease.
XX XX
OS Homo sapiens.
XX XX
PN WO200055180-A2.
XX XX
PD 21-SEP-2000.
XX XX
PF 08-MAR-2000; 2000WO-US05918.
XX XX
PR 12-MAR-1999; 99US-0124270.
XX XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX XX
PI (ROSE/) ROSEN C A.
XX XX
PI Ruben SM;
XX XX
DR WPI: 2000-587514/55.
XX XX
DR N-PSDB; F18045.
XX XX
PT Lung cancer associated gene sequences, referred to as lung cancer
PT antigens, useful for treatment, prevention, and diagnosis of disorders
PT such as lung cancer -
XX XX
PS Claim 11; Page 996-998; 1425pp; English.
XX XX
CC Polynucleotide sequences F17982 - F18424 encode human lung cancer
CC associated proteins represented in B58106 - B58548. Lung cancer
CC associated proteins and polynucleotide sequences, their agonists, and
CC antagonists may have neuroprotective, cytostatic, cardioactive, and
CC immunomodulatory; muscular active general; vulnerrary; gastrointestinal
CC general; nephrotoxic; antiinfective; gynecological; or antibacterial
CC activity. The invention also includes antibodies specific for the protein
CC or polynucleotide sequences. The lung cancer associated polynucleotide
CC sequences may be used for detection of lung cancer, chromosome
CC identification, as chromosome markers, and for numerous other diagnostic
CC or research purposes. The proteins may be used to treat disorders such as
CC neural, immune, muscular, reproductive, gastrointestinal, pulmonary,
CC cardiovascular, renal, and proliferative disorders. The proteins may also
CC be used in the treatment of wounds and infectious diseases.
CC Polynucleotide sequences F18425 - F18433 and peptide B58549 are used in
CC the course of the invention for the identification and characterisation
CC of the polynucleotide and protein sequences.
XX XX
SQ Sequence 406 AA.

Query Match 82.3%; Score 321; DB 21; Length 406;
Best Local Similarity 77.5%; Pred. No. 6,2e-34;
Matches 55; Conservative 9; Mismatches 7; Indels 0; Gaps 0;
QY 1 ORAHGDICTGSGCLQRFITMPFLFCNVNDVCNFAASRNDYSYWLSPALPMNMNAPTGR 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

Db 217 erahbgddlgtagscrlrkstmpflcnlnvncfnasrindyswlstpepmnmapltge 276
 QY 61 ALEPYISRCTV 71
 : 1:111111
 Db 277 nlrpfisrcav 287

QY 61 ALEPYISRCTV 71
 : 1:111111
 Db 100 nlrpfisrcav 110

Search completed: April 13, 2001, 15:41:19
 Job time: 302 sec

RESULT 15
 P93524
 ID P93524 standard; protein; 229 AA.

XX P93524;

DT 04-JUN-1990 (first entry)

DE Complete sequence of the alpha-1-NC1 domain of type IV collagen.

XX Alpha-1-NC1 domain; type IV collagen; cell adhesion;

KW heparin; aortic endothelial cells; metastatic carcinoma M4 cells;

KK rat fibroblasts; MM fibrosarcoma cells; C6 glioma cell;

XX A431 breast carcinoma cells; wound healing; implant acceptance.

XX Key Location/Qualifiers

FT Peptide 201..216

FT /note="TS-1"

FT 49..60

FT /note="TS-2"

FT 17..27

FT /note="TS-3"

XX W08903392-A.

XX 20-APR-1989.

XX 30-AUG-1988; 88WO-US03023.

XX 08-OCT-1987; 87US-0106858.

XX (MINU) MINNESOTA UNIVERSITY.

XX Tsilibary EC, Furcht LT;

XX WPI, 1989-130015/17.

XX Polypeptide(s) with type IV collagen activity -

XX used to promote wound healing, implant acceptance and cellular

XX attachment and inhibit malignant cells

XX Fig 2; page 1/12; 40pp; English.

XX The peptides in the features table are claimed (Claim 1, p. 22). They

XX were synthesised using the Merrifield solid phase method. Binding assays

XX were carried out using peptides TS-1, TS-2 and TS-3. TS-1 promotes

XX adhesions of aortic endothelial cells, metastatic carcinoma M4 cells,

XX normal rat fibroblasts, MM fibrosarcoma cells, C6 glioma cells and

XX A431 breast carcinoma cells. TS-2 binds to type IV collagen, to

XX heparin and promotes adhesion of the above cells. Peptides TS-1, TS-2

XX and TS-3 may be used to promote wound healing and implant acceptance,

XX promote cellular attachment to culture substrata or inhibit the

XX metastasis of malignant cells. They may be used to coat a prosthetic

XX device.

XX Sequence 229 AA;

XX Query Match 80.8%; Score 315; DB 10; Length 229;

XX Best Local Similarity 76.1%; Pred. No. 1,9e-33;

XX Matches 54; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

QY 1 ORAHGQDGLTGLSCQRFETTPFLFCNVNDVCFNFSRNDYSYWLSTPALMPNMAPITGR 60

Db 40 erahbgddlgtagscrlrkstmpflcnlnvncfnasrindyswlstpepmnmaplsgd 99

THIS PAGE BLANK (uspro)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2001, 15:45:37 ; Search time 54.98 Seconds

(without alignments)
151.360 Million cell updates/sec

Title: US-09-543-371-10_COPY_54_124

Perfect score: 390
Sequence: 1 QRHGQDLGTLGSCLOQFTT.....MNAPIGRALPEYISRCTV 71

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues
number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_15:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp.invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp.todent:*
13: sp.unclassified:*
14: sp.vertebrate:*
14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	390	100.0	245	4	Q9NYC4
2	387	99.2	212	4	Q28512
3	374	95.9	203	6	Q28682
4	374	95.9	203	6	Q29032
5	374	95.9	212	6	Q28567
6	372	95.4	210	6	Q28273
7	371	95.1	246	11	Q61435
8	371	95.1	1669	11	Q90250
9	359	92.1	230	11	Q63122
10	325	83.3	979	13	Q919K3
11	321	82.3	229	4	Q9NYC5
12	319	81.8	179	11	P70165
13	319	81.8	253	11	Q61436
14	314	80.5	225	6	Q28271
15	314	80.5	886	4	Q9NDB7
16	311	79.7	161	11	Q61430
17	294	75.4	1752	5	Q07265
18	257	65.9	1758	5	Q19098
19	257	65.9	1759	5	Q19099

20	254	65.1	1747	5	Q26640
21	249	63.8	1802	5	Q17163
22	242	62.1	673	4	Q14052
23	241	61.8	202	6	Q28272
24	235	60.3	1600	4	Q9UEH6
25	235	60.3	1690	4	Q12823
26	235	60.3	1690	4	Q9UMG6
27	235	60.3	1691	4	Q9Y4L4
28	233	59.7	1761	5	Q18407
29	233	59.7	1940	5	Q9YMW5
30	232	59.5	205	6	Q28274
31	219	56.2	208	6	Q29468
32	213	54.6	312	11	Q64457
33	213	54.6	1682	11	Q902R9
34	202	51.8	1779	5	Q9YMW4
35	166	42.6	854	5	Q09238
36	72.5	18.6	106	6	Q9TS99
37	71	18.2	986	5	Q76122
38	64.5	16.5	986	5	Q76803
39	62.5	16.0	252	4	Q15157
40	62.5	16.0	254	4	Q9Y2N0
41	61.5	15.8	470	11	Q05704
42	61.5	15.8	1290	5	Q906L1
43	61	15.6	495	5	Q9Y9R4
44	60.5	15.5	223	8	Q9TC08
45	60.5	15.5	736	6	Q18805

ALIGNMENTS

RESULT	ID	Q9NYC4	PRELIMINARY;	PRT;	245 AA.
AC	Q9NYC4	01-OCT-2000 (TREMBLREL. 15, Created)			
DT	01-OCT-2000 (TREMBLREL. 15, Last sequence update)				
DT	01-OCT-2000 (TREMBLREL. 15, Last annotation update)				
DE	TUMSTATIN (FRAGMENT).				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Maeshima Y., Colorado P.C., Torre A., Holthaus K.A., Grunkemeyer J.A.,				
RA	Erichsen M.D., Hopfer H., Xiao Y., Stillman I.E., Kalluri R.,				
RT	"Distinct anti-tumor properties of a type IV collagen domain derived				
RT	from basement membrane."				
RL	J. Biol. Chem. 0:0-0(2000).				
DR	EMBL; AF238351; AAF72632.1; -				
FT	NON_TER				
FT	SEQUENCE				
SQ	245 AA; 26952 MW; 1EE5028354D9A57D CRC64;				

Query Match 100.0%; Score 390; DB 4; Length 245;
Best Local Similarity 100.0%; Pred. No. 3.8e-42;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	QRHGQDLGTLGSCLOQFTT	PELFCNVNDVCNPFASRNDYSYMLSPALPMNNAPIITGR	60
DB	55	QRHGQDLGTLGSCLOQFTT	PELFCNVNDVCNPFASRNDYSYMLSPALPMNNAPIITGR	114
QY	61	ALPEYISRCTV	71	
DB	115	ALPEYISRCTV	125	
RESULT	2			
ID	Q28512	PRELIMINARY;	PRT;	212 AA.
AC	Q28512;			
DT	01-NOV-1996 (TREMBLREL. 01, Created)			

```
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).
GN COL4A3.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RENAL CORTEX.
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: L47280; AAA91861.1; -.
DR INTERPRO: IPR000504; -.
DR INTERPRO: IPR001442; -.
DR PPRM: PFO1413; C4; 2.
DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 212
SQ SEQUENCE 212 AA; 23469 MW; 4BC574A64E357EE64 CRC64;

Query Match
Best Local Similarity 99.2%; Score 387; DB 6; Length 212;
Matches 70; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORAHGODLGTGSCLORTTPEFLFCNNVDCNFASRNDYSWLTSPALMPMNAIPITGR 60
Db 22 ERASHGODLGTGSCLORTTPEFLFCNNVDCNFASRNDYSWLTSPALMPMNAIPITGR 81
QY 61 ALEPYSRCTV 71
Db 82 ALEPYSRCTV 92

RESULT 3
Q28682 PRELIMINARY; PRT; 203 AA.
AC 028682;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).
GN COL4A3.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RENAL CORTEX.
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: L47283; AAA91893.1; -.
DR INTERPRO: IPR000504; -.
DR INTERPRO: IPR001442; -.
DR PPRM: PFO1413; C4; 2.
DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 203
SQ SEQUENCE 203 AA; 22213 MW; E14173816E4D9E30 CRC64;

Query Match
Best Local Similarity 95.9%; Score 374; DB 6; Length 203;
Matches 66; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORAHGODLGTGSCLORTTPEFLFCNNVDCNFASRNDYSWLTSPALMPMNAIPITGR 60
Db 22 ERASHGODLGTGSCLORTTPEFLFCNNVDCNFASRNDYSWLTSPALMPMNAIPITGR 81
QY 61 ALEPYSRCTV 71
Db 82 ALEPYSRCTV 92

RESULT 4
Q29032 PRELIMINARY; PRT; 203 AA.
AC 029032;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).
GN COL4A3.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RENAL CORTEX.
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: L47284; AAA91882.1; -.
DR INTERPRO: IPR000504; -.
DR INTERPRO: IPR001442; -.
DR PPRM: PFO1413; C4; 2.
DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 203
SQ SEQUENCE 203 AA; 22326 MW; E3B20E32D9A245AA CRC64;

Query Match
Best Local Similarity 93.0%; Score 374; DB 6; Length 203;
Matches 66; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORAHGODLGTGSCLORTTPEFLFCNNVDCNFASRNDYSWLTSPALMPMNAIPITGR 60
Db 22 ERASHGODLGTGSCLORTTPEFLFCNNVDCNFASRNDYSWLTSPALMPMNAIPITGR 81
QY 61 ALEPYSRCTV 71
Db 82 ALEPYSRCTV 92

RESULT 5
Q28567 PRELIMINARY; PRT; 212 AA.
AC 028567;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).
GN COL4A3.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RENAL CORTEX.
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: L47282; AAA91904.1; -.
DR INTERPRO: IPR000504; -.
DR INTERPRO: IPR001442; -.
DR PPRM: PFO1413; C4; 2.
```

```
Db 22 EQAHGODLGTGSCLORTTPEFLFCNNVDCNFASRNDYSWLTSPALMPMNAIPITGR 81
QY 61 ALEPYSRCTV 71
Db 82 ALEPYSRCTV 92

RESULT 4
Q29032 PRELIMINARY; PRT; 203 AA.
AC 029032;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).
GN COL4A3.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RENAL CORTEX.
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: L47284; AAA91882.1; -.
DR INTERPRO: IPR000504; -.
DR INTERPRO: IPR001442; -.
DR PPRM: PFO1413; C4; 2.
DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 203
SQ SEQUENCE 203 AA; 22326 MW; E3B20E32D9A245AA CRC64;

Query Match
Best Local Similarity 95.9%; Score 374; DB 6; Length 203;
Matches 66; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORAHGODLGTGSCLORTTPEFLFCNNVDCNFASRNDYSWLTSPALMPMNAIPITGR 60
Db 22 ERASHGODLGTGSCLORTTPEFLFCNNVDCNFASRNDYSWLTSPALMPMNAIPITGR 81
QY 61 ALEPYSRCTV 71
Db 82 ALEPYSRCTV 92

RESULT 5
Q28567 PRELIMINARY; PRT; 212 AA.
AC 028567;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).
GN COL4A3.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RENAL CORTEX.
RA Turner A.N., Ryan J.J., Derry C.J., Cashman S.J., Katbanna I.,
RA Mason P.J., Pusey C.D.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: L47282; AAA91904.1; -.
DR INTERPRO: IPR000504; -.
DR INTERPRO: IPR001442; -.
DR PPRM: PFO1413; C4; 2.
```


DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 212
 SQ SEQUENCE 212 AA; 23417 MW; 0F5839FCB81BDD8C CRC64;

Query Match 95.9%; Score 374; DB 6; Length 212;
 Best Local Similarity 93.0%; Pred. No. 3.6e-40;
 Matches 66; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORAHGODLCTLSCLOFTTTPMFLCNDVNCNFSRNDYSYWLSTPALPMNMAPIGR 60
 DB 22 EQAHGODLCTLSCLOFTTTPMFLCNDVNCNFSRNDYSYWLSTPALPMNMAPIGR 81

OY 61 ALEPYSRCTV 71
 DB 82 ALEPYSRCTV 92

RESULT 6
 AC 028273 PRELIMINARY; PRT: 210 AA.

DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE COLLAGEN TYPE IV ALPHA 3 CHAIN (FRAGMENT).
 GN COL4A3.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SAVOYED;
 RA MEDLINE=96278820; PubMed=8662866;
 RX Thoner P.S., Zheng K., Kalluri R., Jacobs R., Hudson B.G.;
 RT "Coordinate gene expression of the alpha3, alpha4, and alpha5 chains
 RT of collagen type IV. Evidence from a canine model of X-linked
 RT nephritis with a COL4A5 gene mutation."
 RL J. Biol. Chem. 271:13821-13828(1996).
 DR EMBL: U50935; AAC48585.1; -;
 DR INTERPRO: IPR000504; -;
 DR INTERPRO: IPR001442; -;
 DR PFAM: PF01413; C4; 2;
 DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
 FT NON_TER 1
 FT NON_TER 210
 FT SEQUENCE 210 AA; 23025 MW; 31119EACAB23633D CRC64;

Query Match 95.4%; Score 372; DB 6; Length 210;
 Best Local Similarity 93.0%; Pred. No. 6.5e-40;
 Matches 66; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORAHGODLCTLSCLOFTTTPMFLCNDVNCNFSRNDYSYWLSTPALPMNMAPIGR 60
 DB 32 EQAHGODLCTLSCLOFTTTPMFLCNDVNCNFSRNDYSYWLSTPALPMNMAPIGR 91

OY 61 ALEPYSRCTV 71
 DB 92 ALEPYSRCTV 102

RESULT 7
 AC 061435 PRELIMINARY; PRT: 246 AA.

DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE PROCOLLAGEN TYPE IV, ALPHA 3 (COLLAGEN IV ALPHA 3 CHAIN) (FRAGMENT).
 GN COL4A3.

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALE/C;
 RX MEDLINE=95050957; PubMed=7962065;
 RA Miner J.H., Sanes J.R.;
 RT "Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal
 RT laminae: sequence, distribution, association with laminins, and
 RT developmental switches."
 RL J. Cell Biol. 127:879-891(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALE/C;
 RA Miner J.H.;
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALE/C;
 RA Miner J.H.;
 RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL: Z51566; CAA84529.1; -;
 DR MGD: MGI:104688; Col4a3.
 DR INTERPRO: IPR000504; -;
 DR INTERPRO: IPR001442; -;
 DR PFAM: PF01413; C4; 2;
 DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
 FT NON_TER 1
 FT SEQUENCE 246 AA; 26993 MW; A9B5434F5836F32A CRC64;

Query Match 95.1%; Score 371; DB 11; Length 246;
 Best Local Similarity 93.0%; Pred. No. 1e-39;
 Matches 66; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORAHGODLCTLSCLOFTTTPMFLCNDVNCNFSRNDYSYWLSTPALPMNMAPIGR 60
 DB 56 KRAHGODLCTLSCLOFTTTPMFLCNDVNCNFSRNDYSYWLSTPALPMNMAPIGR 115

OY 61 ALEPYSRCTV 71
 DB 116 ALEPYSRCTV 126

RESULT 8
 AC 090250 PRELIMINARY; PRT: 1669 AA.

DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE ALPHA 3 COLLAGEN IV.
 GN COL4A3.

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=KIDNEY;
 RX MEDLINE=20005934; PubMed=10534397;
 RA Lu W., Phillips C.L., Killen P.D., Haing T., Harrison W.R.,
 RA Elder F.F.B., Miner J.H., Overbeek P.A., Meister M.H.;
 RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a
 RT mouse model of alport syndrome."
 RL Genomics 61:113-124(1999).
 DR EMBL: AF169387; AAD50449.1; -;
 DR MGD: MGI:104688; Col4a3.
 DR INTERPRO: IPR000087; -;
 DR INTERPRO: IPR000504; -;
 DR INTERPRO: IPR001442; -;

PFAM: PF01413; C4; 2.
DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
KM Collagen.
SQ SEQUENCE 1669 AA; 161769 MW; 30976E59739A47B2 CRC64;

Query Match
Best Local Similarity 95.1%; Score 371; DB 11; Length 1669;
Matches 66; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

OY 1 ORAHGODLGTGSCLOFTTPELFCNVNDVCFNFSRNDYSYWLSTPALMPNNAPITGR 60
DB 1479 KRAHGDGLTGTSCLOFTTPELFCNVNDVCFNFSRNDYSYWLSTPALMPNNAPITGR 1538
OY 61 ALEPYISRCTV 71
DB 1539 ALEPYISRCTV 1549

RESULT 9
ID 063122 PRELIMINARY; PRT; 230 AA.
AC 063122;

DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE ALPHA-3 TYPE IV COLLAGEN (FRAGMENT).
GN COL4A3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;

RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=KIDNEY CORTEX;
RA Turner N., Ryan J.J., Derry C.J., Cashman S.J., Katbama I.,
RA Mason P.J., Pusey C.D.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: L47281; AAB72238.2; -.

DR INTERPRO: IPR000504; -
DR INTERPRO: IPR001442; -
DR PFAM: PF01413; C4; 2.
DR PROSITE: PS00030; RNP_1; UNKNOWN_1.
FT NON TER 1
FT NON TER 230
SQ SEQUENCE 230 AA; 25398 MW; 29549E25314CC056 CRC64;

Query Match
Best Local Similarity 92.1%; Score 359; DB 11; Length 230;
Matches 64; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 ORAHGODLGTGSCLOFTTPELFCNVNDVCFNFSRNDYSYWLSTPALMPNNAPITGR 60
DB 40 EHAHGDGLTGTSCLOFTTPELFCNVNDVCFNFSRNDYSYWLSTPALMPNNAPITGR 99
OY 61 ALEPYISRCTV 71
DB 100 ALEPYISRCTV 110

RESULT 10
ID 0919K3 PRELIMINARY; PRT; 979 AA.
AC 0919K3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE *COLLAGEN IV AI CHAIN (FRAGMENT).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.

OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Halter W.M., Dong S.;
RT "Composition, synthesis and assembly of the embryonic chick retinal
RT basal lamina."
RL Dev. Biol. 0:0-0(2000).
DR EMBL: AF239838; AAF44681.1; -.
FT NON TER 1
SQ SEQUENCE 979 AA; 95020 MW; 5B1017D911ED4299 CRC64;

Query Match
Best Local Similarity 83.3%; Score 325; DB 13; Length 979;
Matches 55; Conservative 11; Mismatches 5; Indels 0; Gaps 0;

OY 1 ORAHGODLGTGSCLOFTTPELFCNVNDVCFNFSRNDYSYWLSTPALMPNNAPITGR 60
DB 790 ERAHGDGLTGTSCLOFTTPELFCNVNDVCFNFSRNDYSYWLSTPALMPNNAPITGR 849
OY 61 ALEPYISRCTV 71
DB 850 STRPISRCTV 860

RESULT 11
ID 09NYC5 PRELIMINARY; PRT; 229 AA.
AC 09NYC5.

DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE ARRESTEN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.
RA Colorado P.C., Torre A., Kamphaus G.D., Maeshima Y., Hopfer H.,
RA Takahashi K., Volk R., Zamborsky E.D., Herman S., Sakat P.K.,
RA Erickson M.B., Dhanabal M., Simons M., Post M., Kufe D.,
RA Weichselbaum R.R., Sukhame V.P., Kalluri R.;
RT "Anti-angiogenic cues from vascular basement membrane collagen."
RL Cancer Res. 0:0-0(2000).
DR EMBL: AF258349; AAF72630.1; -.

FT NON TER 1
SQ SEQUENCE 229 AA; 25331 MW; 9693CD100A5CID5 CRC64;

Query Match
Best Local Similarity 82.3%; Score 321; DB 4; Length 229;
Matches 55; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

OY 1 ORAHGODLGTGSCLOFTTPELFCNVNDVCFNFSRNDYSYWLSTPALMPNNAPITGR 60
DB 40 ERAHGDGLTGTSCLOFTTPELFCNVNDVCFNFSRNDYSYWLSTPALMPNNAPITGR 99
OY 61 ALEPYISRCTV 71
DB 100 NIREPISRCTV 110

RESULT 12
ID P70165 PRELIMINARY; PRT; 179 AA.
AC P70165;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE COLLAGEN TYPE IV ALPHAS CHAIN (FRAGMENT).
GN COL4A5.
OS Mus musculus (Mouse).

[illegible]

	Query Match	80.5%	Score 314	DB 4:	Length 886;
	Best Local Similarity	73.2%	Pred No. 7	6e-32;	
	Matches 52;	Conservative 13;	Mismatches 6;	Indels 0;	Gaps 0.
QY	1 ORAHGDDLTGSCIOREFTTTPPLFCNVNDVCFASFASRNDYSWLSLSTPALPMNNATIGR	60			
	:				
DB	697 KRAHGDDLTASSCUREFTTTFMFCNINNOVCNFASRNDYSWLSLSTPEPMPSPMLDKO	756			

OY 61 ALEPYISRCTV 71
:::|:|:|:|
Db 757 SIQPFISRCAY 767

Search completed: April 13, 2001, 15:45:38
Job time: 316 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2001, 15:46:04 ; Search time 20.87 seconds

(without alignments)
116.538 Million cell updates/sec

Title: US-09-543-371-10_COPY_54_124

Perfect score: 390
Sequence: 1 ORAHGDDCTLGSCLOQRTT.....MNMVPIITGRALPEYISRCTV 71

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	390	100.0	1670	1 CA34_HUMAN	Q01955 homo sapien
2	371	95.1	1670	1 CA34_BOVIN	Q28084 bos taurus
3	321	82.3	1669	1 CA14_HUMAN	P02462 homo sapien
4	315	80.8	1669	1 CA14_MOUSE	P02463 mus musculus
5	314	80.5	1669	1 CA54_CANFA	O28247 canis famli
6	314	80.5	1685	1 CA54_HUMAN	P29400 homo sapien
7	256	65.6	1763	1 CA24_ASCSU	P27393 ascaris suu
8	253	64.9	1758	1 CA24_CAEBL	P17140 caenorhabdi
9	252	64.6	1758	1 CA14_CAEBL	P17139 caenorhabdi
10	244	62.6	1707	1 CA24_MOUSE	P08122 mus musculu
11	242	62.1	1712	1 CA24_HUMAN	P08572 homo sapien
12	235	60.3	1678	1 CA64_HUMAN	Q14031 homo sapien
13	222	56.9	1690	1 CA44_HUMAN	P53420 homo sapien
14	219	56.2	1690	1 CA44_BOVIN	Q29442 bos taurus
15	214	54.9	1623	1 CA44_RABIT	P55787 oryctolagus
16	186.5	47.8	1775	1 CA14_DROME	P08120 drosophila
17	63	16.2	628	1 ABFA_ASPNG	Q42254 aspergillus
18	62.5	16.0	622	1 YAK8_SCHRO	O09920 schizosacch
19	62.5	16.0	990	1 ANPM_MANSE	Q11001 manduca sex
20	60	15.4	2517	1 NCR2_HUMAN	O9Y618 h nuclear r
21	59.5	15.3	357	1 CMO_PYRFU	O51741 pyrococcus
22	59	15.1	2210	1 RBP7_EBOSM	O68802 ebola virus
23	58	14.9	397	1 YMP2_YEAST	Q04359 saccharomyc
24	57.5	14.7	265	1 YPR5_ECOLI	P22994 escherichia
25	57.5	14.7	484	1 PANE_HAELN	P44963 haemophilus
26	57.5	14.7	599	1 GP63_LEICH	P15706 leishmania
27	57.5	14.7	810	1 PLMN_ERIEU	Q29485 etinacueus e
28	57.5	14.7	1202	1 RPM2_YEAST	O02773 saccharomyc
29	57	14.6	397	1 UTR2_HUMAN	O15849 homo sapien
30	57	14.6	397	1 UTR2_RABIT	Q28614 oryctolagus
31	57	14.6	610	1 LKHA_CAVPO	P15602 cavia porce
32	57	14.6	1064	1 SYTC_SCHPO	O13651 schizosacch
33	56	14.4	609	1 LKHA_RAT	P30349 rattus norv

ALIGNMENTS

RESULT 1	ID	CA34_HUMAN	STANDARD:	PRT: 1670 AA.	
AC	001955	CA34_HUMAN	STANDARD:	PRT: 1670 AA.	
DT	01-OCT-1996	(Rel. 34, Created)			
DT	15-JUL-1999	(Rel. 38, Last sequence update)			
DT	01-OCT-2000	(Rel. 40, Last annotation update)			
DE	COLLAGEN ALPHA 3(IV) CHAIN PRECURSOR.				
GN	COL4A3.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISUE=Kidney;				
RA	MEDLINE=94364994; PubMed=8083201;				
RA	Mariyama M., Leinonen A., Mochizuki T., Tryggvason K., Reiders S.T.;				
RT	"Complete primary structure of the human alpha 3(IV) collagen chain.				
RT	Coeexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in				
RT	human tissues.";				
RL	J. Biol. Chem. 269:23013-23017(1994).				
RN	[2]				
RP	REVISED.				
RA	Leinonen A.;				
RL	Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.				
RN	[3]				
RP	SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.				
RA	MEDLINE=93015826; PubMed=1400291;				
RA	Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;				
RT	"Exon/intron structure of the human alpha 3(IV) gene encompassing the				
RT	Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially				
RT	antigenic region at the triple helix/NC1 domain junction.";				
RL	J. Biol. Chem. 267:19780-19784(1992).				
RN	[4]				
RP	SEQUENCE OF 1453-1670 FROM N.A.				
RA	MEDLINE=91353570; PubMed=1882840;				
RA	Morrison K.E., Mariyama M., Yang-Peng T.L., Reiders S.T.;				
RT	"Sequence and localization of a partial cDNA encoding the human alpha				
RT	3 chain of type IV collagen.";				
RL	Am. J. Hum. Genet. 49:545-554(1991).				
RN	[5]				
RP	SEQUENCE OF 1331-1670 FROM N.A.				
RC	TISUE=Kidney;				
RA	MEDLINE=92147878; PubMed=1737849;				
RA	Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,				
RA	Pusey C.D.;				
RT	"Molecular cloning of the human Goodpasture antigen demonstrates it				
RT	to be the alpha 3 chain of type IV collagen.";				
RL	J. Clin. Invest. 89:592-601(1992).				
RN	[6]				
RP	SEQUENCE OF 1644-1670 FROM N.A.				
RC	TISUE=Kidney;				
RA	Ding J.;				
RL	Submitted (JAN-1993) to the EMBL/Genbank/DBJ databases.				
RN	[7]				

34	55.5	14.2	666	1 COL4_BRAJA	P98057 bradyrhizob
35	55	14.1	107	1 URE2_ECOLI	O03283 escherichia
36	55	14.1	300	1 ERF5_KRATH	O08341 arabidopsis
37	55	14.1	389	1 UTR1_HUMAN	Q13336 homo sapien
38	55	14.1	929	1 B3AT_MOUSE	P04919 mus musculu
39	55	14.1	3898	1 POLG_HGVA	P19712 hog cholera
40	54.5	14.0	520	1 CCB4_HUMAN	O00305 homo sapien
41	54	13.8	397	1 UTR2_RAT	O62668 rattus norv
42	54	13.8	610	1 LKHA_MOUSE	P24527 mus musculu
43	53.5	13.7	776	1 POLG_LANVY	P29838 langat viru
44	53.5	13.7	1237	1 B3A2_MOUSE	P13808 mus musculu
45	53.5	13.7	1237	1 B3A2_RABIT	P48746 oryctolagus

RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.
RC TISSUE-Kidney:
RX MEDLINE=94124597; Pubmed=82944492;
RA Feng L., Xia Y., Wilson C.B.;
RT "Alternative splicing of the NC1 domain of the human alpha 3(IV)
collagen gene. Differential expression of mRNA transcripts that
predict three protein variants with distinct carboxyl regions";
RL J. Biol. Chem. 269:2342-2348(1994).
RN (8)
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=96196854; Pubmed=9537506;
RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioaka H.,
RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and
alpha(IV) collagen chains are arranged head-to-head on chromosome
2q36";
RL FEBS Lett. 424:11-16(1998).
RN (9)
RP ALTERNATIVE SPLICING.
RX MEDLINE=93280184; Pubmed=8505332;
RA Bernal D., Quiñones S., Saus J.;
RT "The human mRNA encoding the Goodpasture antigen is alternatively
spliced";
RL J. Biol. Chem. 268:12090-12094(1993).
RN (10)
RP VARIANT PRO-1474.
RX MEDLINE=95078827; Pubmed=7987301;
RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,
RA Barrientos A., Monens L.A.H., van Oost B.A., Brunner H.G.,
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal
recessive Alport syndrome";
RL Hum. Mol. Genet. 3:1269-1273(1994).
CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
MESHWORK TOGETHER WITH LAMININS, PROTEGLYCANS AND ENACTIN/
NIDOGEN.
CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOPFORMS, ALPHA 1(IV)-
ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC -1- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOPFORMS: 1 (SHOWN HERE), 2/4 AND
3/15; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER IN THEIR
C-TERMINAL NC1 DOMAINS.
CC -1- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
COCHLEA, LUNG AND BRAIN.
CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
TRIPLE-HELICAL 7S DOMAIN.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPLEPEPTIDE REPEATING
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- PTM: THE ALTERNATIVE SPLICED FORM V CONTAINS AN ADDITIONAL
N-LINKED GLYCOSYLATION SITE.
CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
IV COLLAGENS.
CC -1- DISEASE: ANTIBODIES AGAINST THE NC1 DOMAIN OF ALPHA3(IV) MEDIATE
THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS
CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.
CC -1- DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE I
AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY
GLOMERULONEPHROPATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,
HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN
MALES AND FEMALES.
CC *****
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

[illegible]

```

Query Match      100.0%; Score 390; DB 1; Length 1670;
Best Local Similarity 100.0%; Pred. No. 8e-39;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps
0;

1 ORANGDGLTGGCLOQRTTTPPLPCNVNDVCFNSRNDYSYMLSTPALPMNAPITGR 60
|||||

```

DB 1480. QRAHGDLGLTGLSCLOQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALPMNMAPITGR 1539
 OY 61 ALEPYISRCTV 71
 DB 1540 ALEPYISRCTV 1550

RESULT 2
 CA34_BOVIN STANDARD: PRT: 471 AA.
 AC 028084;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE COLLAGEN ALPHA 3(IV) CHAIN (FRAGMENT).
 GN COLA3.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 NCBI_TaxID=9913;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lens;
 RX MEDLINE=91093146; PubMed=1985905;
 RA Morrison K.E., Germino G.G., Reeders S.T.;
 RT "Use of the polymerase chain reaction to clone and sequence a cDNA
 RT encoding the bovine alpha 3 chain of type IV collagen.";
 RL J. Biol. Chem. 266:34-39(1991).
 RN [12]
 RP SEQUENCE OF 227-258.
 RC TISSUE=Kidney;
 RX MEDLINE=90202779; PubMed=2318822;
 RA Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
 RT "Glomerular basement membrane. Identification of a fourth chain,
 RT alpha 4, of type IV collagen.";
 RL J. Biol. Chem. 265:5466-5469(1990).
 RN [13]
 RP SEQUENCE OF 227-254.
 RX MEDLINE=88330844; PubMed=3417661;
 RA Saus J., Wieslander J., Langeveld J.P.M., Quinones S., Hudson B.G.;
 RT "Identification of the Goodpasture antigen as the alpha 3(IV) chain
 RT of collagen IV.";
 RL J. Biol. Chem. 263:13374-13380(1988).
 RN [14]
 RP SEQUENCE OF 227-244.
 RX MEDLINE=87222419; PubMed=2438283;
 RA Butkowski R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
 RA Hudson B.G.;
 RT "Localization of the Goodpasture epitope to a novel chain of basement
 RT membrane collagen.";
 RL J. Biol. Chem. 262:7874-7877(1987).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
 CC NIDDOGEN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPLET REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M63139; AAA62708.1; -
 DR InterPro: IPR000087; -
 DR InterPro: IPR001442; -
 DR Pfam: PF01413; C4; 2.
 DR Pfam: PF01391; Collagen; 4.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
 FT NON_TER 1 1
 FT DOMAIN <1 238 TRIPLE-HELICAL REGION.
 FT SITE 239 471 NONHELICAL REGION (NC1).
 FT SITE 106 108 CELL ATTACHMENT SITE (POTENTIAL).
 FT MOD_RES 232 232 HYDROXYLATION.
 FT MOD_RES 238 238 HYDROXYLATION.
 FT DISULFID 261 352 OR 349 (BY SIMILARITY).
 FT DISULFID 294 349 OR 352 (BY SIMILARITY).
 FT DISULFID 306 312 BY SIMILARITY.
 FT DISULFID 371 466 OR 463 (BY SIMILARITY).
 FT DISULFID 405 463 OR 466 (BY SIMILARITY).
 FT DISULFID 417 423 BY SIMILARITY.
 FT CONFLICT 253 253 S -> Y (IN REF. 3).
 SQ SEQUENCE 471 AA; 47585 MW; C03B66F14E7008DE CRC64;

Query Match 95.1%; Score 371; DB 1; Length 471;
 Best Local Similarity 91.5%; Pred. No. 4.1e-37;
 Matches 65; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

OY 1 QRAHGDLGLTGLSCLOQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALPMNMAPITGR 60
 DB 281 EQAHGDLGLTGLSCLOQRTTTPFLFCNVNDVCFNFSRNDYSYWLSTPALPMNMAPITGR 340
 OY 61 ALEPYISRCTV 71
 DB 341 ALEPYISRCTV 351

RESULT 3
 CA14_HUMAN STANDARD: PRT: 1669 AA.
 ID CA14_HUMAN
 AC P02462;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE COLLAGEN ALPHA 1(IV) CHAIN PRECURSOR.
 GN COL4A1.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89340433; PubMed=2701944;
 RA Soininen R., Huotari M., Ganguly A., Prockop D.J., Tryggvason K.;
 RT "Structural organization of the gene for the alpha 1 chain of human
 RT type IV collagen.";
 RL J. Biol. Chem. 264:13565-13571(1989).
 RN [12]
 RP SEQUENCE OF 46-1257 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=88083584; PubMed=3691802;
 RA Soininen R., Haka-Risku T., Prockop D.J., Tryggvason K.;
 RT "Complete primary structure of the alpha 1-chain of human basement
 RT membrane (type IV) collagen.";
 RL FEBS Lett. 225:188-194(1987).
 RN [13]


```

FT DOMAIN 1441 1669 NONMEDICAL REGION (NC1).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .).
FT DISULFID 1460 1551 OR 1548.
FT DISULFID 1493 1548 OR 1551.
FT DISULFID 1505 1511 OR 1662.
FT DISULFID 1570 1665 OR 1662.
FT DISULFID 1604 1662 OR 1665.
FT DISULFID 1616 1622
FT CONFLICT 237 238
FT CONFLICT 241 241 SG -> KE (IN REF. 4).
FT CONFLICT 319 319 G -> A (IN REF. 4).
FT CONFLICT 319 319 Q -> A (IN REF. 3).
FT CONFLICT 719 719 N -> D (IN REF. 5).
FT CONFLICT 837 837 D -> Y (IN REF. 5).
FT CONFLICT 842 842 K -> P (IN REF. 5).
FT CONFLICT 896 896 V -> W (IN REF. 5).
FT CONFLICT 904 904 E -> Q (IN REF. 5).
FT CONFLICT 914 914 S -> K (IN REF. 5).
FT CONFLICT 998 998 S -> K (IN REF. 5).
FT CONFLICT 1010 1010 K -> P (IN REF. 5).
FT CONFLICT 1012 1012 S -> K (IN REF. 5).
FT CONFLICT 1358 1358 E -> Q (IN REF. 5).
SEQUENCE 1669 AA: 160611 MW: 388BA6DFB9B8A84 CRC64:

```

```

Query Match 82.3%; Score 321; DB 1; Length 1669;
Best Local Similarity 77.5%; Pred. No. 1.5e-30;
Matches 55; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

```

```

OY 1 ORAHGODLGLTSLCLOREFTTPELFCNVNDVCFASNDYSYWLSTPALMPNNAPIGR 60
    :||||||| :|||||||:|||||||:|||||||:|||||||
Db 1480 ERAHGODLGTAGSCLRKRFSTMPLEFCINNVNCFASNDYSYWLSTPEPMMSMAPIGE 1539

```

```

OY 61 ALEPYISRCTV 71
    :|||||||
Db 1540 NIREPISRCV 1550

```

```

RESULT 4
CA14_MOUSE STANDARD; PRT: 1669 AA.
ID CA14_MOUSE
AC P02463;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE COLLAGEN ALPHA 1(IV) CHAIN PRECURSOR.
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197932; PubMed=2703490;
RA Muthukumar G., Blumberg B., Kurkinen M.;
RT "The complete primary structure for the alpha 1-chain of mouse
collagen IV. Differential evolution of collagen IV domains.";
RL J. Biol. Chem. 264:6310-6317(1989).
RN [2]
RP SEQUENCE OF 1-1154 FROM N.A.
RX MEDLINE=8811221; PubMed=3338566;
RA Wood L., Theriault N., Vogel G.;
RT "cDNA clones completing the nucleotide and derived amino acid
sequence of the alpha 1 chain of basement membrane (type IV) collagen
from mouse.";
RL FEBS Lett. 227:5-8(1988).
RN [3]
RP SEQUENCE OF 1149-1424 FROM N.A.
RX MEDLINE=86301886; PubMed=3755692;
RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogel G.;
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
synthetic oligodeoxynucleotide.";
RL Gene 43:301-304(1986).
RN [4]

```

```

RP SEQUENCE OF 1276-1669 FROM N.A.
RX MEDLINE=85127033; PubMed=2578961;
RA Oberbeumer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
RA Vogel G., Voss T., Siebold B., Glanville R.W., Kuhn K.;
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
the alpha 1(IV) chain of basement membrane collagen as derived from
complementary DNA.";
RL Eur. J. Biochem. 147:217-224(1985).
RN [5]
RP SEQUENCE OF 1441-1669 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
alpha 1(IV) and alpha 2(IV) collagen.";
RL J. Biol. Chem. 263:8496-8499(1987).
RN [6]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86196099; PubMed=3009468;
RA Sakurai Y., Sullivan M., Yamada Y.;
RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
collagen genes.";
RL J. Biol. Chem. 261:6654-6657(1986).
RN [7]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogel G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Barbelo P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
RN [9]
RP SEQUENCE OF 1-129 FROM N.A.
RX MEDLINE=88243724; PubMed=3379041;
RA Killen P.D., Barbelo P., Sakurai Y., Yamada Y.;
RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
collagen chain and the corresponding region of the gene.";
RL J. Biol. Chem. 263:8706-8709(1988).
CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
GLomerular BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
MESHWORK TOGETHER WITH LAMININS, PROTEGLYCANS AND ENACTIN/
NIDOGEN.
CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) -
ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
TRIPLE-HELICAL 7S DOMAIN.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPLET REPEATING
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
IV COLLAGENS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: J03758; AAA37439.1; -
DR EMBL: M23333; AAA51625.1; -
DR EMBL: J04694; AAA50292.1; -

```

```

RC STRAIN=SAMOVED: TISSUE=Kidney:
RA MEDLINE=94224868; PubMed=8171024;
RA Zheng K., Thoner P.S., Matrano P., Bauml R., McInnes R.R.:
RT "Canine x-chromosome-linked hereditary nephritis: a genetic model for
RT human x-linked hereditary nephritis resulting from a single base
RT mutation in the gene encoding the alpha 5 chain of collagen type
RT IV."
RL Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).
CC -I- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC MESHWORK TOGETHER WITH LAMININS, PROTEOLYANS AND ENACTIN/
CC NIDOGEN.
CC -I- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) -
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -I- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC -I- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -I- PTR: PROLINES ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -I- PTR: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTR- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -I- DISEASE: A DEFECT IN COL4A5 HAS BEEN FOUND TO BE THE CAUSE OF
CC CANINE X-LINKED HEREDITARY NEPHRITIS (HN), A DISEASE SIMILAR TO
CC THAT IN HUMANS (ALSO REFERRED TO AS ALPORT SYNDROME) CHARACTERIZED
CC BY PROGRESSIVE RENAL FAILURE AND NEUROSENSORY DEAFNESS.
CC -I- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/sib.ch).
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U07888; AAB60258.1; -.
CC DR HSSP; 004656; 2AM0.
CC DR InterPro; IPR0000087; -.
CC DR InterPro; IPR001442; -.
CC DR Pfam; PF01413; C4; 2.
CC DR Pfam; PF01391; Collagen; 8.
CC KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
CC Glycoprotein; Basement membrane; Collagen; Cell adhesion.
CC MW TER 1 1
CC NON_TER 1 1
CC DOMAIN <1 530 TRIPLE-HELICAL REGION.
CC FT 531 >754 NONHELICAL REGION (NC1).
CC FT DISULFID 552 643 OR 640 (BY SIMILARITY).
CC FT DISULFID 585 640 OR 643 (BY SIMILARITY).
CC FT DISULFID 597 603 BY SIMILARITY.
CC FT DISULFID 662 ? OR 754 (BY SIMILARITY).
CC FT DISULFID 696 754 BY SIMILARITY.
CC FT DISULFID 708 714 BY SIMILARITY.
CC FT NON_TER 754 754 BY SIMILARITY.
CC SEQUENCE 754 AA; 73537 MW; D5E321C287FA925B CRC64;

Query Match 80.5%; Score 314; DB 1; Length 754;
Best Local Similarity 73.2%; Pred. No. 4,6e-30;
Matches 52; Conservative 13; Mismatches 6; Gaps 0; Indels 0;

1 ORAAGDGTGAGCGAORTTMTPECCANVDVCFNSRDYSTWSTPALMNNAPITGR 60
Db 572 KRAAGDGTGAGCGAORTTMTPECCANVDVCFNSRDYSTWSTPALMNNAPITGR 631
61 ALPEYISRCIV 71
632 SIQFISRCIV 642

```

RESULT 6
 CAS4_HUMAN STANDARD; PRT; 1685 AA.
 ID CAS4_HUMAN
 AC P29400; Q16126; Q16006;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE COLLAGEN ALPHA 5(IV) CHAIN PRECURSOR.
 GN COL4A5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 ON NCB1_TaxID=9606;
 RX MEDLINE-94165049; PubMed-8120014;
 RA Zhou J., Leinonen A., Tryggvason K.;
 RL J. Biol. Chem. 269:6608-6614(1994).
 [2]
 RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.
 RC TISSUE-Kidney;
 RX MEDLINE-92316923; PubMed-1352287;
 RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
 RT "Complete amino acid sequence of the human alpha 5 (IV) collagen
 chain and identification of a single-base mutation in exon 23
 converting glycine 521 in the collagenous domain to cysteine in an
 Alport syndrome patient.";
 RL J. Biol. Chem. 267:12475-12481(1992).
 [3]
 RP SEQUENCE OF 85-1685 FROM N.A.
 RC TISSUE-Placenta;
 RX MEDLINE-90337990; PubMed-2380186;
 RA Pihlajaniemi T., Pihlajaniemi E.R., Myers J.C.;
 RT "Complete primary structure of the triple-helical region and the
 carboxyl-terminal domain of a new type IV collagen chain, alpha
 5(IV)."
 RL J. Biol. Chem. 265:13758-13766(1990).
 [4]
 RP SEQUENCE OF 924-1685 FROM N.A.
 RX MEDLINE-91169491; PubMed-2004755;
 RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
 RT "Characterization of the 3' half of the human type IV collagen alpha
 5 gene that is affected in the Alport syndrome.";
 RL Genomics 9:1-9(1991).
 [5]
 RP SEQUENCE OF 914-1685 FROM N.A.
 RX MEDLINE-90160375; PubMed-1689491;
 RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B.,
 Tryggvason K.;
 RT "Identification of a distinct type IV collagen alpha chain with
 restricted kidney distribution and assignment of its gene to the
 locus of X chromosome-linked Alport syndrome.";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).
 [6]
 RP SEQUENCE OF 1442-1471 FROM N.A.
 RX MEDLINE-90352791; PubMed-233699;
 RA Myers J.C., Jones T.A., Pihlajaniemi E.R., Kadri A.S., Goddard A.D.,
 Sheer D., Solomon E., Pihlajaniemi T.;
 RT "Molecular cloning of alpha 5(IV) collagen and assignment of the gene
 to the region of the X chromosome containing the Alport syndrome
 locus.";
 RL Am. J. Hum. Genet. 46:1024-1033(1990).
 [7]
 RP SEQUENCE OF 1-20 FROM N.A.
 RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.,
 Marynen P.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 [8]
 RP SEQUENCE OF 1258-1270 FROM N.A. (SPLICED FORM).
 RX MEDLINE-94133540; PubMed-8301933;
 RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H.,
 Cassiman J.-J., Marynen P.;
 RT "Differential splicing of COL4A5 mRNA in kidney and white blood
 cells: a complex mutation in the COL4A5 gene of an Alport patient
 deletes the NCI domain.";
 RL Kidney Int. 44:1316-1321(1993).
 [9]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-97338662; PubMed-9195222;
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
 RT "The clinical spectrum of type IV collagen mutations.";
 RL Hum. Mutat. 9:477-499(1997).
 [10]
 RP VARIANT AS SER-1564.
 RX MEDLINE-91169492; PubMed-1672282;
 RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L.,
 Tryggvason K.;
 RT "Single base mutation in alpha 5(IV) collagen chain gene converting a
 conserved cysteine to serine in Alport syndrome.";
 RL Genomics 9:10-18(1991).
 [11]
 RP VARIANT AS ARG-325.
 RX MEDLINE-92303559; PubMed-1376965;
 RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Grunfeld J.-P.,
 Tryggvason K., Gubler M.-C., Antignac C.;
 RT "Substitution of arginine for glycine 325 in the collagen alpha 5
 (IV) chain associated with X-linked Alport syndrome: characterization
 of the mutation by direct sequencing of PCR-amplified lymphoblast
 cDNA fragments.";
 RL Am. J. Hum. Genet. 51:135-142(1992).
 [12]
 RP VARIANT AS GLU-325.
 RX MEDLINE-93244772; PubMed-1363780;
 RA Renieri A., Serri M., Myers J.C., Pihlajaniemi T., Massella L.,
 Rizzone G., de Marchi M.;
 RT "De novo mutation in the COL4A5 gene converting glycine 325 to
 glutamic acid in Alport syndrome.";
 RL Hum. Mol. Genet. 1:127-129(1992).
 [13]
 RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
 RX MEDLINE-94010948; PubMed-8406498;
 RA Lemmink H.H., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J.,
 Tryggvason K., Haggma-Schouten W.A.G., Roodvoets A.P., Rascher W.,
 van Oost B.A., Smeets H.J.M.;
 RT "Identification of four novel mutations in the COL4A5 gene of
 patients with Alport syndrome.";
 RL Genomics 17:485-489(1993).
 [14]
 RP VARIANTS AS E-400; V-406; V-638; A-638; R-653; R-796; R-869; R-872 & C-1241.
 RX MEDLINE-95322976; PubMed-7599631;
 RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;
 RT "Detection of 12 novel mutations in the collagenous domain of the
 COL4A5 gene in Alport syndrome patients.";
 RL Hum. Mutat. 5:197-204(1995).
 [15]
 RP VARIANT AS ARG-1649.
 RX MEDLINE-96213750; PubMed-8651292;
 RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
 Denison J.C., Fain P.R., Gregory M.C.;
 RT "A mutation causing Alport syndrome with tardive hearing loss is
 common in the western United States.";
 RL Am. J. Hum. Genet. 58:1157-1165(1996).
 [16]
 RP VARIANTS AS.
 RX MEDLINE-96213754; PubMed-8651296;
 RA Renieri A., Brutini M., Galli L., Zanelli P., Neri T., Rossetti S.,
 Turco A., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,
 Scolari F., Sessa A., Rizzoni G., Tryggvason K., Pignatti P.F.,
 Savi M., Ballabio A., de Marchi M.;
 RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
 exons of the COL4A5 gene.";
 RL Am. J. Hum. Genet. 58:1192-1204(1996).
 [17]

RP VARIANT AS ASP-1498.
 RX MEDLINE=96233932; PubMed=8829632.
 RA Tverskaya S., Bobrylna V., Tsalykova F., Ignatova M.,
 RA Krasnopolskaya X., Evgrafov O.;
 RT "Substitution of A1498D in noncollagen domain of $\alpha 1(\text{IV})$ collagen
 RT chain associated with adult-onset X-linked Alport syndrome.";
 RL Hum. Mutat. 7:149-150(1996).
 RL [18]
 RP VARIANT AS GLN-1677.
 RX MEDLINE=97295089; PubMed=9150741;
 RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
 RT "Common ancestry of three Ashkenazi-American families with Alport
 RT syndrome and COL4A5 R1677Q.";
 RL Hum. Genet. 99:681-684(1997).
 RL [19]
 RP VARIANTS AS R-174; R-177; R-325; C-1410; W-1421; T-1517 AND D-1596.
 RX MEDLINE=98112435; PubMed=9452056;
 RA Neel T.M., Zaneli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
 RA Pignatti G.F., Galli L., Bruttini M., Renieri A., Mingarelli R.,
 RA Trivelli A., Pinciaroli A.R., Ragaiolo M., Rizzoni G.F., de Marchi M.;
 RT "Missense mutations in the COL4A5 gene in patients with X-linked
 RT Alport syndrome.";
 RL Hum. Mutat. Suppl. 1:S106-S109(1998).
 RL [20]
 RP VARIANTS AS, AND VARIANTS.
 RX MEDLINE=99140256; PubMed=10094548;
 RA Plant K.E., Green P.M., Vetric D., Pinter F.A.;
 RT "Detection of mutations in COL4A5 in patients with Alport syndrome.";
 RL Hum. Mutat. 13:124-132(1999).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCAN AND EXTACTIN/
 CC NIDOGEM.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1 (SHOWN HERE) AND 2; ARE
 CC PROBABLY PRODUCED BY ALTERNATIVE SPLICING. ISOFORM 2 IS A LONGER
 CC ISOFORM FOUND IN KIDNEY, IN WHICH 2 EXTRA G-X-X REPEATS INTO THE
 CC TRIPLE-HELIX DOMAIN ARE INTRODUCED.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -1- DISEASE: DEFECTS IN COL4A5 ARE ASSOCIATED WITH X-LINKED ALPORT
 CC SYNDROME (AS). ALPORT SYNDROME IS CHARACTERIZED WITH PROGRESSIVE
 CC GLOMERULONEPHRITIS, OFTEN ASSOCIATED WITH HIGH-TONE SENSORY/NEURAL
 CC DEAFNESS, SPECIFIC EYE ABNORMALITIES (LENTICONS AND MACULAR
 CC FLECKS), AND GLOMERULAR BASEMENT MEMBRANE DEFECTS. IN MALES, THE
 CC TYPICAL TIME COURSE FOR THE PROGRESS OF ALPORT SYNDROME IS:
 CC HEMATURIA BY THE AGE OF 5 YEARS, DEAFNESS AND HYPERTENSION IN

Query Match 80.58; Score 314; DB 1; Length 1665;
 Best Local Similarity 73.28; Pred. No. 1e-29;
 Matches 52; Conservative 13; Mismatches 6; Indels 0; Gaps 0;

QY 1 ORANGODLGTGSGLOFTMPFLFCVNVNCFASRNDYSYMLTPALPMNAPTGR 60
 DB 1496 KRAAGDGLTGRGSLRFRSTMPFECNINNVCFASRNDYSYMLSTPEPMDMGLPKQ 1555

QY 61 ALPEYISRCTV 71
 DB 1556 SIQPFISRCAY 1566

RESULT 7

CA24.ASCSU
 ID CA24.ASCSU STANDARD; PRT; 1763 AA.
 AC P27393;
 DT 01-ANG-1992 (Rel. 23, Created)
 DT 01-ANG-1992 (Rel. 23, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE COLLAGEN ALPHA 2(IV) CHAIN PRECURSOR.
 OS Ascaris sum (Pig roundworm) (Ascaris lumbricoides).
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascarididae; Ascarididae;
 NCBI_TaxID=6253;
 RX MEDLINE=91340768; PubMed=1714907;
 RA Pettit J., Kingston I.B.;
 RT "The complete primary structure of a nematode alpha 2(IV) collagen
 RT and the partial structural organization of its gene.";
 RL J. Biol. Chem. 266:16149-16156(1991).
 CC -1- SUBUNIT: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.
 CC -1- FUNCTION: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.
 CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH
 CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1
 CC DOMAINS.
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: I (SHOWN HERE) AND II; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M67507; AAA18014.1;
 DR PIR: S16366; S16366.
 DR InterPro: IPR000087;
 DR InterPro: IPR001442;
 DR Pfam: PF01413; C4.2.
 DR Pfam: PF01391; Collagen; 23.
 KW Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;
 KW Alternative splicing; Glycoprotein; Signal.
 FT SIGNAL 1 26
 FT CHAIN 27 1763
 FT DOMAIN 27 42
 FT DOMAIN 43 1529
 FT DOMAIN 1530 1763
 FT DISULFID 1548 1637
 FT DISULFID 1581 1634
 FT DISULFID 1593 1599
 FT DISULFID 1656 1752
 FT DISULFID 1680 1749
 FT DISULFID 1702 1709
 FT CARBOHYD 126 126
 FT CARBOHYD 249 249
 FT VARSPLIC 230 266
 FT SEQUENCE 1763 AA; 168526 MW; 304F528BC06AAED CRC64;
 FT ISOFORM II)

CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; 222964; CAA80536.1; -;
DR	EMBL; 222964; CAA80537.1; -;
DR	EMBL; J05066; AAA27989.1; -;
DR	EMBL; U23227; AAA64312.1; ALT-SEQ.
DR	PIR; A34476; A34476.
DR	InterPro: IPR000087; -;
DR	InterPro: IPR001442; -;
DR	Pfam; PF01391; Collagen; 23.
DR	Pfam; PF01391; Collagen; 23.
KW	Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;
KW	Alternative splicing; Glycoprotein; Signal.
FT	SIGNAL
FT	1
FT	26
FT	POTENTIAL.
FT	COLLAGEN ALPHA 2(IV) CHAIN.
FT	75 DOMAIN.
FT	DOMAIN
FT	42
FT	1527
FT	TRIPLE-HELICAL REGION.
FT	NONHELICAL REGION (NC1).
FT	OR 1632 (BY SIMILARITY).
FT	OR 1632 (BY SIMILARITY).
FT	OR 1632 (BY SIMILARITY).
FT	BY SIMILARITY.
FT	BY SIMILARITY.
FT	OR 1747 (BY SIMILARITY).
FT	OR 1747 (BY SIMILARITY).
FT	OR 1750 (BY SIMILARITY).
FT	BY SIMILARITY.
FT	O-LINKED (GLYCOSAMINOGLYCAN) (POTENTIAL).
FT	GLDSVGRPGRGPRGSGSSLYGPRGNPEKGDK -> G
FT	DIGMAGPAGPGPPIASTWSKTTIGPKDGLGEKGEK (IN
FT	ISOFORM 11).
FT	G -> E (IN MN114; 73% LETHAL).
FT	G -> T (IN MN126; 100% LETHAL).
FT	G -> E (IN MN109; 37% LETHAL).
FT	G -> R (IN MN103 AND MN151; 96% LETHAL).
FT	G -> R (IN MN152; 50% LETHAL).
FT	G -> R (IN MN101; 100% LETHAL).
FT	G -> E (IN MN129; 100% LETHAL).
FT	G -> E (IN MN143; 100% LETHAL).
FT	G -> R (IN G30; 90% LETHAL).
FT	G -> R (IN E1470; 94% LETHAL).
FT	G -> E (IN MN139; 20% LETHAL).
FT	G -> D (IN G25; 2% LETHAL).
FT	G -> D (IN MN147; 7% LETHAL).
FT	G -> D (IN G37 AND B246; 9% LETHAL).
FT	P -> L (IN REF. 1).
FT	CONFLICT
FT	1682
FT	1682
FT	SEQUENCE
FT	1758 AA; 167750 MW; 97EE3F5DBB2D2AC5 CRC64;
QY	1 ORAHGODIGTSGCLOQRTTMEFLFCNVNDVCNEASRNDYSWLSTPALMPNNAPITGR 60
QY	::: ::: ::: ::: ::: ::: ::: :::
DB	1566 EKSHNODIGHAGSCLOQRSTMFLECDENNVCNYSASRREKSYWLSTSPAIP--MAPVNER 1623
QY	61 ALEPIRSCTV 71
QY	::: ::: ::: ::: ::: ::: ::: :::
DB	1624 EIEPIISRCAY 1634
RESULT	9
CA14-CAEEL	STANDARD; PRT; 1758 AA.
AC	P17139;
DT	01-AUG-1990 (Rel. 15, Created)
DT	01-NOV-1992 (Rel. 21, Last sequence update)
DT	01-NOV-1997 (Rel. 35, Last annotation update)
DE	COLLAGEN ALPHA 1(IV) CHAIN PRECURSOR.
GN	EMB-2 OR CLB-2 OR KO4H4.1.
GN	Cacenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
 CC Rhabditidae; Peioderinae; Caenorhabditis.
 CX NCBI_TaxID=6239;
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=91141582; PubMed=1996137;
 RA Guo X., Johnson J., Kramer J.M.;
 RT "Embryonic lethality caused by mutations in basement membrane
 collagen of C. elegans.";
 RL Nature 349:707-709(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Pavello A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Kershaw J., Kirsten T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten T., Laister N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
 RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Stalton J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Watson J., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Woldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 RN [3]
 RP SEQUENCE OF 1446-1758 FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=90008929; PubMed=2793871;
 RA Guo X., Kramer J.M.;
 RT "The two Caenorhabditis elegans basement membrane (type IV) collagen
 genes are located on separate chromosomes.";
 RL J. Biol. Chem. 264:17574-17582(1989).
 CC -1- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANS.
 CC -1- SUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.
 CC TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH
 CC INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NC1
 CC DOMAINS.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
 CC TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
 CC IV COLLAGENS.
 CC -1- DISEASE: MUTATIONS IN THIS GENE CAUSE TEMPERATURE-SENSITIVE
 CC LETHALITY DURING LATE EMBRYOGENESIS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X56979; CAA40299.1; -;
 DR EMBL: Z27078; CAA81584.1; -;
 DR EMBL: J05067; AAB59179.1; -;
 DR PIR: B34476; B34476;
 DR PIR: S13651; S13651;
 DR Wormpep: K04H4.1; CE00246.
 DR InterPro: IPR000087; -;
 DR InterPro: IPR001442; -;
 DR Pfam: PF01413; C4; 2.

DR Pfam: PF01391; Collagen; 22.
 KW Extracellular matrix; Basement membrane;
 KW Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 ? 7194 ?
 FT PROPEP 1 ? 7194 ?
 FT CHAIN 2195 1758
 FT DOMAIN 195 1529
 FT DOMAIN 1530 1758
 FT DISULFID 1549 1640
 FT DISULFID 1582 1637
 FT DISULFID 1594 1600
 FT DISULFID 1659 1754
 FT DISULFID 1693 1751
 FT DISULFID 1705 1711
 FT VARIANT 402 402
 FT VARIANT 408 408
 FT CONFLICT 1514 1514
 FT CONFLICT 1514 185
 FT CONFLICT 259 263
 FT CONFLICT 302 304
 FT CONFLICT 366 404
 FT CONFLICT 581 581
 FT CONFLICT 768 768
 FT CONFLICT 813 814
 FT CONFLICT 830 830
 FT CONFLICT 1275 1275
 FT CONFLICT 1722 1722
 FT SEQUENCE 1738 AA; 170857 MW; 7083D9AF3E05D45 CRC64;
 Query Match 64.6%; Score 252; DB 1; Length 1758;
 Best Local Similarity 64.3%; Pred. No. 3e-22;
 Matches 45; Conservative 8; Mismatches 17; Indels 0; Gaps 0;
 QY 2 RAHQDGLTLCSCQRTPTPLFCNNVNDQNFNRDYSWTLPLMPMNAPIGTA 61
 DB 1570 RASGDDLPQSCSLSKNTMFCNNMNSVCHVSSRDYSEWLTDEPMTPMNPVTGTA 1629
 QY 62 LEPTISRCTV 71
 DB 1630 IRPTISRCV 1639
 RESULT 10
 CA24_MOUSE STANDARD; PRT; 1707 AA.
 ID CA24_MOUSE 061375;
 AC 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE COLLAGEN ALPHA 2(IV) CHAIN PRECURSOR.
 GN COL4A2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_TaxID=10090;
 OX [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89197933; PubMed=2703491;
 RA Saus J., Quinones S., Mackrell A., Blumberg B., Muthukumar G.,
 RA Pihlajaniemi T., Kurkinen M.;
 RT "The complete primary structure of mouse alpha 2(IV) collagen.
 Alignment with mouse alpha 1(IV) collagen.";
 RL J. Biol. Chem. 264:6318-6324(1989).
 RN [2]
 RP SEQUENCE OF 1-33 FROM N.A.
 RX MEDLINE=89066738; PubMed=3198626;
 RA Kayles P., Wood L., Theriault N., Kurkinen M., Vogel G.;
 RT "Head-to-head arrangement of murine type IV collagen genes.";
 RL J. Biol. Chem. 263:19274-19277(1988).

RN [3]
 RP SEQUENCE OF 970-1480 FROM N.A.
 RX MEDLINE=86220192; PubMed=3011432;
 RA Schwarz U., Schuppan D., Oberbauer I., Glanville R.W.,
 RT Deutmann R., Timpi R., Kuehn K.;
 RT "Structure of mouse type IV collagen. Amino-acid sequence of the C-terminal 511-residue-long triple-helical segment of the alpha 2(IV) chain and its comparison with the alpha 1(IV) chain.";
 RL Eur. J. Biochem. 157:49-56(1986).
 RN [4]
 RP SEQUENCE OF 1480-1707 FROM N.A.
 RX MEDLINE=87054581; PubMed=3780963;
 RA Schwarz-Magdalen U., Oberbauer I., Kuehn K.;
 RT "CDNA and protein sequence of the NCI domain of the alpha 2-chain of collagen IV and its comparison with alpha 1(IV).";
 RL FEBS Lett. 208:203-207(1986).
 RN [5]
 RP SEQUENCE OF 1481-1707 FROM N.A.
 RX MEDLINE=87250460; PubMed=3597383;
 RA Kurknen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
 RT Saus J., Pihlajaniemi T.;
 RT "Extensive homology between the carboxyl-terminal peptides of mouse alpha 1(IV) and alpha 2(IV) collagen.";
 RL J. Biol. Chem. 262:8496-8499(1987).
 RN [6]
 RP SEQUENCE OF 1041-1489 FROM N.A.
 RX MEDLINE=87005245; PubMed=3758345;
 RA Vogel G., Horn E., Carter J., Kayes P.S.;
 RT "Proposed alignment of helical interruptions in the two subunits of the basement membrane (type IV) collagen.";
 RL FEBS Lett. 206:29-32(1986).
 RN [7]
 RP SEQUENCE OF 964-1003; 1005-1085 AND 1087-1109 FROM N.A.
 RX MEDLINE=85296379; PubMed=3839906;
 RA Kurknen M., Bernard M.P., Barlow D.P., Chow L.T.;
 RT "Characterization of 64-, 123- and 182-base-pair exons in the mouse alpha 2(IV) collagen gene.";
 RL Nature 317:177-179(1985).
 RN [8]
 RP SEQUENCE OF 1-60 FROM N.A.
 RX MEDLINE=89071159; PubMed=3200851;
 RA Burdello P.D., Martin G.R., Yamada Y.;
 RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE' MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/NIIDGEN.
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOMERFORMS, ALPHA 1(IV)-ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL TRIPLE-HELICAL 7S DOMAIN.
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
 CC -1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M23334; AAA51626.1; -

DR EMBL; M23333; AAA51626.1; JOINED.
 DR EMBL; J04695; AAA50293.1; -
 DR EMBL; J04448; AAA37438.1; -
 DR EMBL; X04647; CAA28308.1; -
 DR EMBL; X04647; CAA28308.1; -
 DR EMBL; M15833; AAA37341.1; -
 DR EMBL; X04410; CAA27998.1; -
 DR EMBL; X02896; CAA26655.1; -
 DR EMBL; X02897; CAA51614.1; -
 DR EMBL; X02898; CAA26657.1; -
 DR EMBL; X02899; CAA26658.1; -
 DR EMBL; M23334; AAA51627.1; -
 DR PIR; A33526; A33526.
 DR HSSP; P19972; 1KVD.
 DR MGD; MGI:88455; Col4a2.
 DR InterPro; IPR000087; -
 DR InterPro; IPR001442; -
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 19.
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Glycoprotein; Basement membrane; Collagen; Signal.
 FT SIGNAL 1 25
 FT PROPEP 26 183
 FT CHAIN 184 1707
 FT DOMAIN 184 1479
 FT DOMAIN 1480 1707
 FT DISULFID 1499 1588
 FT DISULFID 1532 1585
 FT DISULFID 1544 1550
 FT DISULFID 1607 1703
 FT DISULFID 1641 1700
 FT DISULFID 1653 1660
 FT CARBOHYD 138 138
 FT CARBOHYD 1270 1270
 FT CONFLICT 1051 1051
 FT CONFLICT 1097 1097
 FT CONFLICT 1171 1171
 FT CONFLICT 1179 1179
 FT CONFLICT 1241 1241
 FT CONFLICT 1328 1328
 FT CONFLICT 1573 1573
 FT CONFLICT 1623 1623
 SQ SEQUENCE 1707 AA; 167391 MW; 1A565159605F508 CRC64;
 Query Match 62.68; Score 244; DB 1; Length 1707;
 Best Local Similarity 62.0%; Pred. No. 2.6e-21;
 Matches 44; Conservative 10; Mismatches 15; Indels 2; Gaps 1;
 QY 1 ORAHGDLTGSGLORETFMPFLICNVNDVCNFRSDYSYMLSTPALMNMNAPITGR 60
 DB 1519 EKAHNODGLAGSLARSTMPFLYCNPGDYCYASRNDKSYLWSTA--PLPMMPVAEE 1576
 QY 61 ALEPYISRCVY 71
 DB 1577 EIKYISRCV 1587
 RESULT 11
 CA24_HUMAN
 ID CA24_HUMAN STANDARD; PRT; 1712 AA.
 AC P08572;
 DT 01-AUG-1988 (Rel. 08, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE COLLAGEN ALPHA 2(IV) CHAIN PRECURSOR.
 GN COL4A2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxId=9606;
 RX [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89066769; PubMed=3198637;

CC	-1-	SUBUNIT: THESE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) -
CC	ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE	
CC	WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.	
CC	-1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS	
CC	DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE	
CC	G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY	
CC	CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL	
CC	TRIPLE-HELICAL 7S DOMAIN.	
CC	-1- PPM: PROLINES AT THE THIRD POSITION OF THE TRIPEPIDE REPEATING	
CC	UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.	
CC	-1- PPM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH	
CC	ARE INVOLVED IN INTER- AND INTRA-MOLECULAR DISULFIDE BONDING. 12 OF	
CC	THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE	
CC	IV COLLAGENS.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; X05562; CAA29076.1; -	
DR	EMBL; X05610; CAA29098.1; -	
DR	EMBL; J02760; AAA58422.1; -	
DR	EMBL; M36963; AAA53099.1; -	
DR	EMBL; X12784; CAA31275.1; -	
DR	EMBL; J04217; AAA53097.1; -	
DR	PIR; A32024; A32024.	
DR	MIM; 120090; -	
DR	InterPro; IPR000087; -	
DR	InterPro; IPR001442; -	
DR	Pfam; PF01413; C4; 2.	
DR	Pfam; PF01391; Collagen; 20.	
KW	Extracellular matrix; Connective tissue; Repeat; Hydroxylation;	
KM	Glycoprotein; Basement membrane; Collagen; Signal.	
FT	SIGNAL 1 25	
FT	PROPEP 26 183	
FT	CHAIN 184 1712	
FT	DOMAIN 184 1484	
FT	DISULFD 1485 1712	
FT	DISULFD 1504 1593	
FT	DISULFD 1537 1590	
FT	DISULFD 1549 1555	
FT	DISULFD 1612 1708	
FT	DISULFD 1646 1705	
FT	DISULFD 1658 1665	
FT	CARBOHYD 138 138	
FT	CONFLICT 471 471	
FT	CONFLICT 683 683	
FT	CONFLICT 1575 1575	
FT	CONFLICT 1663 1663	
FT	CONFLICT 1701 1701	
SEQ	SEQUENCE 1712 AA; 167535 MW; 2582A17847890037 CRC64;	
QY	Query Match 62.1%; Score 242; DB 1; Length 1712; Best Local Similarity 62.0%; Pred. No. 4.5e-21;	
Matches	44; Conservative 10; Mismatches 15; Indels 2; Gaps 1;	
Db	1 ORAHODIGTGSGCRPTMPFEICNVNDVDFNFSRDNDYSWLSTPLMPNNMPTGR 60	
	: : : :	
DB	1524 EKANHDDGLAGSCCLARSITMFELCNFGDVCIYASRNDSKSWLSTTA--PLPMNPVAED 1581	
	:: ::	
QY	61 ALEPYISRCTV 71	
	:: ::	
DB	1582 EIKPYISRCSV 1592	
	:	
RESULT	12	
CA64_HUMAN	STANDARD:	
	PRT; 1678 AA.	

CC		014031.	
DT	01-NOV-1997	(Rel. 35, Created)	
DT	01-NOV-1997	(Rel. 35, Last sequence update)	
DT	01-OCT-2000	(Rel. 40, Last annotation update)	
DE	COLLAGEN ALPHA 6(IV) CHAIN PRECURSOR.		
GN	COLA46.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=EYE, and KIDNEY:		
RX	MEDLINE=94117179; PubMed=8129572;		
RA	Ohashi T., Sugimoto M., Mattei M.-G., Ninomiya Y. ;		
RT	Identification of a new collagen IV chain, alpha 6(IV), by cDNA		
RT	isolation and assignment of the gene to chromosome Xq42, which is the		
RL	same locus for COLA45.*;		
J. Biol. Chem.	269:7520-7526(1994).*		
-I-	FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF		
GLomerular basement membranes (GBM), FORMING A 'CHICKEN-WIRE'			
MESHWORK TOGETHER WITH LAMININS, PROTEOLYCATS AND ENACTIN/ NIDOGEN.			
-I-	SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-		
ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE			
WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.			
-I-	SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).		
-I-	DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS		
DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE			
G-X-Y REPEATS IN THE LONG CENTRAL TRIPEL-HELICAL DOMAIN (WHICH MAY			
CAUSE FLEXIBILITY IN THE TRIPEL HELIX), AND A SHORT N-TERMINAL			
TRIPLE-HELICAL 7S DOMAIN.			
-I-	PM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING		
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.			
-I-	PM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH		
ARE INVOLVED IN INTER- AND INTRA-MOLECULAR DISULFIDE BONDING. 12 OF			
THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE			
IV COLLAGENS.			
-I-	SIMILARITY: TO OTHER TYPE IV COLLAGENS.		

CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
CC	EMBL; D21337; BAA04809.1; .		
DR	MIM; 303631; .		
DR	InterPro: IPR0000087; .		
DR	InterPro: IPR001442; .		
DR	Pfam; PF01413; C4; 2.		
DR	Pfam; PF01391; Collagen; 21.		
KW	Extracellular matrix; Connective tissue; Basement membrane;		
Repeat; Hydroxylation; Glycoprotein; Cell adhesion; Collagen; Signal.			
FT	SIGNAL	1	15
FT	CHAIN	16	1678
FT	DOMAIN	16	45
FT	DOMAIN	46	1450
FT	DOMAIN	1451	1678
FT	SITE	514	516
FT	SITE	559	561
FT	SITE	985	987
FT	DISULFD	1469	1558
FT	DISULFD	1502	1555
FT	DISULFD	1514	1520
FT	DISULFD	1577	1674
FT	DISULFD	1611	1671
FT	DISULFD	1623	1630
FT	CARBOHYD	126	126
SEQUENCE	1678 AA; 162461 MW; 656EBA282D3D37BD CRC64;		

```
Query Match Similarity      60.3%; Score 235; DB 1; Length 1678;
Best Local Similarity       57.7%; Pred. No. 3, le-20;
Matches    41; Conservative   14; Mismatches   14; Indels     2; Gaps        1;
```

```
Oy          1 ORAHQDGLTGSGLCIQRTFTMPFLFCNVNDYCNFASRNDYSWYLSTPALPMNMAPITGR 60
              ::::| | | | | | | | | | : | : | : | : | : | | | | | | | | 1 : | : | : 
Db         1489 EKAHPIIDGFGAGSCLPRESTMPEFYICINIECHVARRDKSWYLSSTA--PIPMPVSQT 1546

Oy          61 ALERPSRCTV 71
               : |||||:|
Db           1547 QIPYISRCRV 1557
```

RESULT_13

ID	CAT4_HUMAN	STANDARD:	PRT:
AC	P53420.		1690 AA.
DJ	01-OCT-1996 (Rel. 34, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DE	01-OCT-2000 (Rel. 40, Last annotation update)		
DR	COLLAGEN ALPHA 4(IV) CHAIN PRECURSOR.		
GN	COL4A4.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. NCBI_TaxId=9606;		
RX	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE-Kidney;		
RA	MEDLINE=95014445; PubMed=7523402;		
RT	"Leinonen A., Marjama M., Mochizuki T., Tryggvason K., Reeders S.T.; chain. Comparison with structure and expression of the other alpha (IV) chains." ;		
RL	J. Biol. Chem. 269:26172-26177(1994).		
RN	[2]		
RP	SEQUENCE OF 1219-1690 FROM N.A.		
RC	TISSUE-Eye;		
RA	MEDLINE=93374047; PubMed=8365481;		
RT	Sugimoto M., Ohashi T., Yoshioaka H., Matsuo N., Ninomiya Y.; cDNA isolation and partial gene structure of the human alpha 4(IV) collagen chain." ;		
RL	Febbs Lett. 330:122-128(1993).		
RN	[3]		
RP	SEQUENCE OE 1407-1507 FROM N.A.		
RA	MEDLINE=93054733; PubMed=1429714;		
RT	Kanagata Y., Matel M.-G., Niimoriya Y.; Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha 4 chain of basement membrane collagen type IV and assignment of the gene to the distal long arm of human chromosome 2." ;		
RL	J. Biol. Chem. 267:23753-23758(1992).		
RN	[4]		
RP	REVIEW ON VARIANTS.		
RA	MEDLINE=97338662; PubMed=9195222;		
RT	Lemlink H.H., Schroeder C.H., Monnens L.A.H., Smets H.J.M.; The clinical spectrum of type IV collagen mutations." ;		
RL	Hum. Mutat. 9:477-499(1997).		
RN	[5]		
RP	VARIANT AS SER-1201.		
RA	MEDLINE=95078927; PubMed=7987396;		
RT	Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Guhlcr M.-C., Pison Y., Verellen-Dumoulin C., Chan B., Schroeder C.H., Smets H.J., Reeders S.T.; Identification of mutations in the alpha 3(IV) and alpha 4(IV) collagen genes in autosomal recessive Alport syndrome." ;		
RL	Nat. Genet. 8:77-82(1994).		
RN	[6]		
RP	VARIANT BRH GIU-897.		
RA	MEDLINE=96379660; PubMed=8787673;		
RT	Lemmink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H., Brunner H.G., van Oost B.A., Monnens L.A.H., Smets H.J.M.; Benign familial hematuria due to mutations of the type IV collagen		


```

CC "The alpha 4(IV) chain of basement membrane collagen. Isolation of
RT cDNAs encoding bovine alpha 4(IV) and comparison with other type IV
RT collagens."
RN J. Biol. Chem. 267:1253-1258(1992).
[2]
RN SEQUENCE OF 217-246.
RP MEDLINE=90202779; PubMed=2318822;
RX Gunwar S., Saus J., Noelken M.E., Hudson B.G.;
RT "Glomerular basement membrane. Identification of a fourth chain,
RT alpha 4, of type IV collagen."
RN J. Biol. Chem. 265:5466-5469(1990).
[3]
RN SEQUENCE OF 217-233.
RX MEDLINE=8722419; PubMed=2438283;
RA Butkowski R.J., Langeveld J.P.M., Wieselander J., Hamilton J.,
RA Hudson B.G.;
RT "Localization of the Goodpasture epitope to a novel chain of basement
RT membrane collagen."
RN J. Biol. Chem. 262:7874-7877(1987).
CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/
CC NIDOEIN.
CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-
CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC -1- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE
CC COLocalIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,
CC COCHLEA, LUNG AND BRAIN.
CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC TRIPLE-HELICAL 7S DOMAIN.
CC -1- PPM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- PPM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC IV COLLAGENS.
CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
-----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announcement/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: M77480; AAA30458.2; ALT_SEQ.
DR InterPro: IPR0000087; -.
DR InterPro: IPR001442; -.
DR Pfam: PF01413; CA; 2.
DR Pfam: PF01391; Collagen; 3.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT NON TER 1 1
FT DOMAIN <1 222 TRIPLE-HELICAL REGION.
FT DOMAIN 223 453 NONHELICAL REGION (NC1).
FT DISULFID 243 453 OR 329 (BY SIMILARITY).
FT DISULFID 243 332 OR 329 (BY SIMILARITY).
FT DISULFID 276 329 OR 332 (BY SIMILARITY).
FT DISULFID 288 294 BY SIMILARITY.
FT DISULFID 351 449 OR 446 (BY SIMILARITY).
FT DISULFID 385 446 OR 449 (BY SIMILARITY).
FT DISULFID 397 404 BY SIMILARITY.
FT DISULFID 219 219 I -> P (IN REF. 2 AND 3).
FT CONFLICT 453 AA; 46384 MW; E7ED10AE9A65BC1 CRC64;
FO SEQUENCE

```

Query Match	56.2%;	Score 219;	DB 1;	Length 453;
Best Local Similarity	53.5%;	Pred. No. 6.6e-19;		

	Matches	38, Conservative	14, Mismatches	17, Indels	2, Gaps	1, Ambiguous
QY	1	ORAHODIGTGSCGQRFETMPELTCNVNDVQNFASRNDYSLWLSPPALMNNMAPITGR	60			
Db	263	ERAHNODDGLAGSCPLPFTSTPFAVCNIHQVCHYARARNDSYWLASTA--PLMPPLUSD	320			
QY	61	ALEPIISRCTV	71			
Db	321	ETRPYISRCVAV	331			

RESULT	15
CA44_RABIT	
ID	CA44_RABIT
AC	P55787
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	30-MAY-2000 (Rel. 39, Last annotation update)
DE	COLLAGEN ALPHA 4(IV) CHAIN (FRAGMENT).
GN	COL4A4.
OS	Oryctolagus cuniculus (Rabbit).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX	NCBI_Taxid=9986;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Corneal endothelium;
RX	MEDLINE=95054733; PubMed=1429714;
RA	Kanagata Y., Mattei M.-G., Ninomiya Y.;
RT	"Isolation and sequencing of cDNAs and genomic DNAs encoding the
RT	alpha 4 chain of basement membrane collagen type IV and assignment of
RL	the gene to the distal long arm of human chromosome 2.";
J.	Biol. Chem. 267:23753-23758(1992).
CC	-1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC	GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'
CC	MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND EXTRACTIN/
CC	NIDOGEN.
CC	-1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) -
CC	ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC	WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.
CC	-1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC	-1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC	DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC	G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC	CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC	TRIPLE-HELICAL 7S DOMAIN.
CC	-1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC	UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC	-1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC	ARE INVOLVED IN INTRA- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC	THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC	IV COLLAGENS.
CC	-1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC	-----
CC	CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	CC the European Bioinformatics Institute. There are no restrictions on its
CC	CC use by non-profit institutions as long as its content is in no way
CC	CC modified and this statement is not removed. Usage by and for commercial
CC	CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	CC or send an email to license@isb-sib.ch).
CC	-----
CC	CC EMBL: L01477; -; NOT_ANNOTATED_CDS.
DR	InterPro: IPR0000087; -;
DR	InterPro: IPR001442; -;
DR	Pfam: PF01413; CA; 2.
DR	Pfam: PF01391; Collagen; 5.
RV	Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW	Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT	NON_TER 1 1
FT	DOMAIN <1 392 TRIPLE-HELICAL REGION.
FT	DOMAIN 493 623 NONHELICAL REGION (NC1).
FT	DISULFID 413 502 OR 499 (BY SIMILARITY).

FT DISULFID 446 499 OR 502 (BY SIMILARITY).
FT DISULFID 458 464 BY SIMILARITY.
FT DISULFID 521 619 OR 616 (BY SIMILARITY).
FT DISULFID 555 616 OR 619 (BY SIMILARITY).
FT DISULFID 567 574 BY SIMILARITY.
SO SEQUENCE 623 AA; 62393 MM; CCBC9BB31242FE82 CRC64;

Query Match 54.9%; Score 214; DB 1; Length 623;
Best Local Similarity 50.7%; Pred. No. 3.7e-18;
Matches 36; Conservative 15; Mismatches 18; Indels 2; Gaps 1;

OY 1 ORAHGODIGTIGSCIQRTTPELFCNVNDYCNFASRNDYSYWLSTPALMPNNAPITGR 60
::||| ||| ||| |::||::||| ||| ||| |||::| |::
DB 433 EKAHNDLGLAGSCLPFTSTLPFAICNIHQYCHVQRNDKSYLASAG--PLPMPPLSEE 490
OY 61 ALEPYISRCTV 71
: ||||| |
DB 491 EIRPYISRCV 501

Search completed: April 13, 2001, 15:46:08
Job time: 281 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2001, 15:42:38 ; Search time 33.96 seconds
(without alignments)
143.679 Million cell updates/sec

Title: US-09-543-371-10_COPY 54_124

Perfect score: 390

Sequence: 1 QRAHGODLGLTGLSCGLRFTT.....MNNAPITGRALPEYISCTV 71

Scoring table: BLOSUM62
Gap10.0, Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database:

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	390	100.0	1670	1 CGH03B	collagen alpha 3(I
2	382	97.9	220	2 B49736	collagen alpha 3(I
3	371	95.1	246	2 148302	collagen alpha 3(I
4	371	95.1	471	2 A39024	collagen alpha 3(I
5	321	82.3	1669	1 CGH04B	collagen alpha 1(I
6	319	81.8	253	2 148304	collagen alpha 5(I
7	315	80.8	1669	1 CGH04B	collagen alpha 1(I
8	314	80.5	754	2 A53267	collagen alpha 5(I
9	314	80.5	1691	1 S22917	collagen alpha 5(I
10	311	79.7	161	2 S49488	collagen alpha 3(I
11	306	78.5	258	2 B61228	collagen alpha 1(I
12	294	75.4	1752	2 A45407	collagen alpha 3(I
13	257	65.9	1758	2 T29350	hypothetical prote
14	257	65.9	1758	2 T29351	collagen alpha 2(I
15	256	65.6	1763	2 S16366	collagen alpha 2(I
16	254	65.1	1747	2 A54121	collagen alpha-4 c
17	253	64.9	261	2 A34476	collagen alpha 2(I
18	252	64.6	1744	2 S40991	collagen alpha 1(I
19	244	62.6	1707	2 A33526	collagen alpha 2(I
20	242	62.1	1712	1 CGH02B	collagen alpha 2(I
21	235	60.3	1691	1 CGH06B	collagen alpha 6(I
22	233	59.7	1761	2 T13990	collagen type IV a
23	231	59.2	775	2 A61228	collagen alpha 2(I
24	222	56.9	1690	1 CGH01B	collagen alpha 4(I
25	219	56.2	453	2 S18804	collagen alpha 4(I
26	214	54.9	623	2 A45137	collagen alpha 4(I
27	213	54.6	312	2 148303	collagen alpha 4(I
28	186.5	47.8	1775	2 A31893	collagen alpha 1(I
29	98	25.1	79	2 C43928	probable collagen

30	78.5	20.1	58	2 B43928
31	62.5	16.0	252	2 JC2399
32	62.5	16.0	622	2 S62580
33	61.5	15.8	470	2 S17447
34	60.5	15.5	736	2 T03849
35	60	15.4	1495	2 S60255
36	59.5	15.3	613	2 S55615
37	59	15.1	555	2 T23531
38	59	15.1	1777	2 T00490
39	58	14.9	397	2 S54018
40	58	14.9	428	2 C70670
41	58	14.9	1445	2 T10728
42	58	14.9	1729	2 T43403
43	57.5	14.7	484	2 H64105
44	57.5	14.7	599	2 B42049
45	57.5	14.7	599	2 A44951

ALIGNMENTS

RESULT 1
CGH03B
collagen alpha 3(IV) chain precursor, long splice form - human
N:Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice for
C:Species: Homo sapiens (man)
C:Date: 28-Oct-1994 #sequence, revision 03-Oct-1995 #text, change 22-Jun-1999
C:Accession: A54763; A43928; A44043; A45971; A39286
R:Matsumoto, M.; Kishimoto, A.; Mochizuki, T.; Tryggvason, K.; Reiders, S.T.
J. Biol. Chem. 269, 23013-23017, 1994
A:Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpress
A:Reference number: A54763; MUID:94364994
A:Accession: A54763
A:Molecule type: mRNA
A:Residues: 1-1670 <MAR>
A:Cross-references: GB:M80031; NID:9577563; PID:9577564
A:Experimental source: kidney
R:Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
J. Clin. Invest. 89, 592-601, 1992
A:Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the
A:Reference number: A43928; MUID:92147878
A:Accession: A43928
A:Molecule type: mRNA
A:Residues: 1331-1524, 'I', 1526-1670 <TUR>
A:Cross-references: GB:M81379
A:Experimental source: kidney
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 267, 19780-19784, 1992
A:Title: Exon/Intron structure of the human alpha 3(IV) gene encompassing the Goodpas
cton.
A:Reference number: A44043; MUID:93015826
A:Accession: A44043
A:Molecule type: DNA; mRNA
A:Residues: 1386-1670 <QUT>
A:Cross-references: GB:M92993; NID:9177895; PID:AAA21610.1; PID:9177896
A:Note: sequence extracted from NCBI backbone (NCBI:115597)
R:Quinones, S.; Bernal, D.; Garcia-Sogo, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 269, 17358, 1994
A:Reference number: A44738; MUID:94274734
A:Contents: annotation; erratum; correction to intronic sequence in A44043
R:Bernal, D.; Quinones, S.; Saus, J.
J. Biol. Chem. 268, 12090-12094, 1993
A:Title: The human mRNA encoding the Goodpasture antigen is alternatively spliced.
A:Reference number: A45971; MUID:93280184
A:Accession: A45971
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1427-1444 <BER>
A:Note: sequence extracted from NCBI backbone (NCBI:133363); sequence incorrectly id
Am. J. Hum. Genet. 49, 545-554, 1991
A:Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain
A:Reference number: A39786; MUID:91353570

A:Accession: A39786
A:Molecule type: mRNA
A:Residues: 1453-1593, 'A', 1595-1670 <MOR>
A:Cross-references: GB:555790; NID:9234418; PIDN:AA19637.1; PID:9234419
C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope
C:Genetics:
A:Gene: GDB:COL4A3
A:Cross-references: GDB:128351; OMIM:120070
A:Map position: 2q36-2q37
A:Insertions: 1385/1; 1418/1; 1468/1; 1547/2; 1585/3; 1643/2 #status incomplete
A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
mong trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric a
er associations in the interrupted helical domain (with disulfide and desmosine cross-l
C:Function:
A:Description: minor structural component of extracellular basement membrane in kidney g
C:Superfamily: collagen alpha 1(IV) chain
Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extracel
F:29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MAT>
F:29-1670/Product: collagen alpha 3(IV) chain, long splice form #status predicted <MAT>
F:43-1438/Region: interrupted helical, NH1 <NH1>
F:791-793/Region: cell attachment (R-G-D) motif
F:996-998/Region: cell attachment (R-G-D) motif
F:1154-1156/Region: cell attachment (R-G-D) motif
F:1306-1308/Region: cell attachment (R-G-D) motif
F:1345-1347/Region: cell attachment (R-G-D) motif
F:1432-1434/Region: cell attachment (R-G-D) motif
F:1433-1434/Region: carboxyl-terminal nonhelical, NCI <NCI>
F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>
F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:31-33, 39, 41, 125, 422, 476, 479, 662, 722, 809, 1387/disulfide bonds: interchain #status predi
F:253/Binding site: carboxylate (asn) (covalent) #status predicted
F:1460-1548, 1493-1551/disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
F:1505-1511, 1616-1622/disulfide bonds: #status predicted
F:1570-1662, 1604-1665/disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 100.0%; Score 390; DB 1; Length 1670;
Best Local Similarity 100.0%; Pred. No. 3.1e-38;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORAHGODLGTGSCLOPRTTTPFLFCNVNDVNCNFSASNDYSYWLSTPALMPMNAPIITGR 60
DB 1480 ORAHGODLGTGSCLOPRTTTPFLFCNVNDVNCNFSASNDYSYWLSTPALMPMNAPIITGR 1539
OY 61 ALEPISRCTV 71
DB 1540 ALEPISRCTV 1550

RESULT 2
B49736
collagen alpha 3(IV) chain, medium splice form - human (fragment)
N:Contains: collagen alpha 3(IV) chain, splice form GP-V
C:Species: Homo sapiens (man)
C:Date: 03-May-1994 #sequence, revision 12-Nov-1999 #text_change 17-Mar-2000
C:Accession: B49736; D49736; S69111
R:Feing, L.; Xia, Y.; Wilson, C.B.
J. Biol. Chem. 269, 2342-2348, 1994
A:Title: Alternative splicing of the NCI domain of the human alpha3(IV) collagen gene.
A:Reference number: A49736; MUID:94124597
A:Accession: B49736
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 169-220 <FEN1>
A:Accession: D49736
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 22-220 <FEN2>
A:Cross-references: GB:U02519; NID:9409106; PIDN:AA18942.1; PID:9409107

A:Note: this is the conceptual translation of the nucleic acid submitted to GenBank
R:Pendres, J.R.; Bernal, D.; Revert, F.; Johansson, C.; Fresquet, V.J.; Cervera, J.;
Eur. J. Biochem. 229, 754-760, 1995
A:Title: Characterization and expression of multiple alternatively spliced transcript
utaminogen and one of its alternative forms.
A:Reference number: S69111; MUID:95278230
A:Accession: S69111
A:Molecule type: mRNA
A:Residues: 1-45, 169-204, 'L', 206-220 <PEN>
C:Comment: For the complete sequence of the long splice form, see PIR:CGH03B.
C:Genetics:
A:Gene: GDB:COL4A3
A:Cross-references: GDB:128351; OMIM:120070
A:Map position: 2q36-2q37
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; extra
F:1-45, 169-220/Product: collagen alpha 3(IV) chain, medium splice form (fragment) #status pr
F:22-220/Product: collagen alpha 3(IV) chain, splice form GP-V (fragment) #stat
F:34-134/Domain: collagen IV carboxyl-terminal repeat <CT1>

Query Match 97.9%; Score 382; DB 2; Length 220;
Best Local Similarity 98.6%; Pred. No. 3.5e-38;
Matches 70; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORAHGODLGTGSCLOPRTTTPFLFCNVNDVNCNFSASNDYSYWLSTPALMPMNAPIITGR 60
DB 63 ORAHGODLGTGSCLOPRTTTPFLFCNVNDVNCNFSASNDYSYWLSTPALMPMNAPIITGR 122
OY 61 ALEPISRCTV 71
DB 123 ALEPISRCTV 133

RESULT 3
I48302
collagen alpha 3(IV) chain - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 15-Mar-1996 #sequence, revision 15-Mar-1996 #text_change 16-Feb-1997
C:Accession: I48302; S47278
R:Miner, J.H.; Sanes, J.R.
J. Cell Biol. 127, 879-891, 1994
A:Title: Collagen IV alpha 3, alpha 4, and alpha 5 chains in rodent basal laminae: se
A:Reference number: A54979; MUID:95050557
A:Accession: I48302
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-246 <RES>
A:Cross-references: EMBL:235166; NID:9535197; PID:9535198
C:Superfamily: collagen alpha 1(IV) chain

Query Match 95.1%; Score 371; DB 2; Length 246;
Best Local Similarity 93.0%; Pred. No. 8.1e-37;
Matches 66; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORAHGODLGTGSCLOPRTTTPFLFCNVNDVNCNFSASNDYSYWLSTPALMPMNAPIITGR 60
DB 56 ORAHGODLGTGSCLOPRTTTPFLFCNVNDVNCNFSASNDYSYWLSTPALMPMNAPIITGR 115
OY 61 ALEPISRCTV 71
DB 116 ALEPISRCTV 126

RESULT 4
A39024
collagen alpha 3(IV) chain - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Dec-1992 #sequence, revision 04-Dec-1992 #text_change 13-Aug-1999
C:Accession: A39024; S20672; S17802; A35167; C39419; S13747; S20815
R:Morrison, K.E.; Germino, G.G.; Reenders, S.T.

J. Biol. Chem. 266, 34-39, 1991
 A>Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the
 A:Reference number: A39024; MUID:91093146
 A:Accession: A39024
 A:Molecule type: mRNA
 A:Residues: 1-471 <MOR>
 A:Cross-references: EMBL:M63139; NID:q162886; PIDD:AAA62708.1; PID:q162887
 R:Burkowski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.
 J. Biol. Chem. 262, 7874-7877, 1987
 A>Title: Localization of the Goodpasture epitope to a novel chain of basement membrane
 A:Reference number: S18432; MUID:87222419
 A:Accession: S20672
 A:Molecule type: protein
 A:Residues: 227-228, 'X', 230-244 <BUT>
 R:Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.
 J. Biol. Chem. 263, 13374-13380, 1988
 A>Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen
 A:Reference number: S17802; MUID:88330844
 A:Accession: S17802
 A:Molecule type: protein
 A:Residues: 227-228, 'X', 230-252, 'Y', 254 <SNU>
 R:Günther, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.
 J. Biol. Chem. 263, 5466-5469, 1990
 A>Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of type
 A:Reference number: A35167; MUID:90202779
 A:Accession: A35167
 A:Molecule type: protein
 A:Residues: 236-258 <GUN>
 R:Günther, S.; Ballestier, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; Noe
 J. Biol. Chem. 266, 15318-15324, 1991
 A>Title: Glomerular basement membrane. Identification of dimeric subunits of the noncoll
 A:Reference number: A39419; MUID:91332055
 A:Accession: C39419
 A:Molecule type: protein
 A:Residues: 236-255 <GUN>
 C:Superfamily: collagen alpha 1(IV) chain
 C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; e
 F.1-238/Domain: collagenous (fragment) #status predicted <COL>
 F.239-471/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NCI>
 F.239-353/Domain: repeat NCI #status predicted <NCI>
 F.354-471/Domain: repeat NCI #status predicted <NCI>
 F.232,233/Modified site: hydroxyproline (Pro) #status experimental
 F.306-312,417-423/Disulfide bonds: #status predicted

Query Match 95.1%; Score 371; DB 2; Length 471;
 Best Local Similarity 91.5%; Pred. No. 1,6e-36;
 Matches 65; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

1 ORAHGDDLTGSCLOFTTPEFLFCNVNDVCFASRNDYSYWLSTPPALPMNAPITGR 60
 :::
 DB 281 EQAHGDDLTGSCLOFTTPEFLFCNVNDVCFASRNDYSYWLSTPPALPMNAPITGR 340

QY 61 ALEPTISRCTV 71
 |||||
 DB 341 ALEPTISRCTV 351

RESULT 5
 CGH04B
 collagen alpha 1(IV) chain precursor - human
 N:Alternate names: procollagen alpha 1(IV) chain
 C:Species: Homo sapiens (man)
 C>Date: 28-May-1986 #sequence_revision 31-Dec-1992 #ext-change 07-Dec-1999
 C:Accession: S16876; A32117; S00238; S00048; S25826; A32115; S00207; S39614; A02863; A58
 R:Solinen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggyvason, K.
 J. Biol. Chem. 264, 13565-13571, 1989
 A>Title: Structural organization of the gene for the alpha-1 chain of human type IV coll
 A:Reference number: S16876; MUID:89340433
 A:Accession: S16876
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1669 <SO11>

A:Cross-references: EMBL:J04217; GB:J05039; NID:q180800; PIDD:AAA53098.1; PID:q180803
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1988
 R:Solinen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggyvason, K.
 J. Biol. Chem. 263, 17217-17220, 1988
 A>Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen
 A:Reference number: A92690; MUID:89034231
 A:Accession: A32117
 A:Molecule type: DNA
 A:Residues: 1-28 <SO12>
 A:Cross-references: EMBL:J04217; NID:q180759; PIDD:AAA53097.1; PID:q553233
 R:Poeschl, E.; Pollner, R.; Kuehn, K.
 EMBO J. 7, 2687-2695, 1988
 A>Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membran
 A:Reference number: S02738; MUID:89030632
 A:Accession: S02738
 A>Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-6, 'L', 8-28 <POE>
 A:Cross-references: EMBL:X12784; NID:q30072
 R:Brazel, D.; Oberbauer, I.; Dieringer, H.; Babel, W.; Glanville, R.W.; Deutzmann,
 Eur. J. Biochem. 168, 529-536, 1987
 A>Title: Completion of the amino acid sequence of the alpha1 chain of human basement
 A:Reference number: S00048; MUID:88029471
 A:Accession: S00048
 A:Molecule type: mRNA
 A:Residues: 1-318, 'A', 320-944 <BRA1>
 A:Cross-references: EMBL:X05561; NID:q30066; PIDD:CAA29075.1; PID:q30067
 A:Accession: S25826
 A:Molecule type: protein
 A:Residues: 271-318, 'A', 320-554 <BRA2>
 R:Glanville, R.W.; Qian, R.Q.; Siebold, B.; Risteli, J.; Kuehn, K.
 Eur. J. Biochem. 152, 213-219, 1985
 A>Title: Amino acid sequence of the N-terminal aggregation and cross-linking region (
 A:Reference number: A23115; MUID:86004708
 A:Accession: A23115
 A:Molecule type: protein
 A:Residues: 28-236, 'KE', 239-240, 'K', 242-243 <GIA>
 A:Experimental source: placenta
 A>Note: the amino end of the mature form is blocked
 R:Solinen, R.; Haka-Risku, T.; Prockop, D.J.; Tryggyvason, K.
 FEBS Lett. 225, 188-194, 1987
 A>Title: Complete primary structure of the alpha(1)-chain of human basement membrane
 A:Reference number: S00207; MUID:88083584
 A:Accession: S00207
 A:Molecule type: mRNA
 A:Residues: 244-530 <SO13>
 A:Cross-references: EMBL:Y00706; NID:q29548; PIDD:CAA68698.1; PID:q29549
 R:Edle, J.A.; Golbl, R.; Mann, K.; Kuehn, K.
 EMBO J. 12, 4795-4802, 1993
 A>Title: The alpha-1-beta-1 integrin recognition site of the basement membrane collag
 A:Reference number: S39614; MUID:94038963
 A:Accession: S39614
 A:Molecule type: protein
 A:Residues: 371-554 <EBL>
 R:Babel, W.; Glanville, R.W.
 Eur. J. Biochem. 143, 545-556, 1984
 A>Title: Structure of human basement-membrane (type IV) collagen. Complete amino-acid
 A:Reference number: A02863; MUID:85003629
 A:Accession: A02863
 A:Molecule type: protein
 A:Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 9
 A:Experimental source: placenta
 R:Glanville, R.W.; Rauter, A.
 Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981
 A>Title: Peptide fragments of human placental basement-membrane collagens showing inte
 A:Reference number: S16908; MUID:82005835
 A:Accession: A58517
 A:Molecule type: protein
 A:Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553, 1389-1405, 'XX', 1408-1409, 'X', 1411
 R:MacKnight, R.S.; Benson, V.A.; Lovellio, K.T.; van der Rest, M.; Fietzek, P.P.
 Biochemistry 22, 4940-4948, 1983
 A>Title: Isolation and characterization of pepsin-solubilized human basement membrane
 A:Reference number: S16910; MUID:84053346

A:Residues: 1-185,'L',187-318,'S',320-368,'L',370-402,'F',404-480,'L',482-492,'H',494-71
A:Cross-references: EMBL:X06777
R:Killem, P.D.; Burbeio, P.; Sakurai, Y.; Yamada, Y.
J. Biol. Chem. 263, 8706-8709, 1988
A:Title: Structure of the amino-terminal portion of the murine alpha-1(IV) collagen chain
A:Reference number: A28066; MUID:88243724
A:Accession: A28066
A:Molecule type: mRNA
A:Residues: 1-129 <K11>
A:Cross-references: EMBL:J03758; NID:q192669; PIDN:AAA37439.1; PID:q192670
R:Oberbauer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogell, G.; Voss,
Eur. J. Biochem. 147, 217-224, 1985
A:Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1
A:Reference number: A02864; MUID:85127033
A:Accession: A02864
A:Molecule type: mRNA
A:Residues: 1276-1669 <OER>
A:Cross-references: EMBL:X02201; NID:q50233; PIDN:CAA26132.1; PID:q1333876
R:Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogell, G.
Gene 43, 301-304, 1986
A:Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeox
A:Reference number: A25636; MUID:86301886
A:Accession: A25636
A:Molecule type: mRNA
A:Residues: 1149-1396,'S',1398-1424 <NAT>
A:Cross-references: EMBL:M14042; NID:q192286; PIDN:AAA37342.1; PID:q192287
A:Note: the authors translated the codon CAG for residue 1374 as Arg
R:Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihla
J. Biol. Chem. 262, 8496-8499, 1987
A:Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)
A:Reference number: A94680; MUID:87250460
A:Accession: A29301
A:Molecule type: mRNA
A:Residues: 1441-1669 <KUR>
A:Cross-references: EMBL:M15832; NID:q192282; PIDN:AAA37340.1; PID:q387115
R:Killem, P.D.; Burbeio, P.D.; Martin, G.R.; Yamada, Y.
J. Biol. Chem. 263, 12310-12314, 1988
A:Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequenc
A:Reference number: S19079; MUID:88315019
A:Accession: S19079
A:Molecule type: DNA
A:Residues: 1-28 <R12>
A:Cross-references: EMBL:J03944; NID:q192673; PIDN:AAA37442.1; PID:q466503
R:Kayes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogell, G.
J. Biol. Chem. 263, 19274-19277, 1988
A:Title: Head-to-head arrangement of murine type IV collagen genes.
A:Reference number: A92702; MUID:89066738
A:Accession: A32003
A:Molecule type: DNA
A:Residues: 1-28 <KAY>
A:Cross-references: EMBL:J04448; NID:q192666; PIDN:AAA37437.1; PID:q450449
R:Burbeio, P.D.; Martin, G.R.; Yamada, Y.
Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988
A:Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional prom
A:Reference number: A94220; MUID:89071759
A:Accession: A31766
A:Molecule type: DNA
A:Residues: 1-28 <BUR>
A:Cross-references: EMBL:M23333; NID:q340878; PIDN:AAA51625.1; PID:q535668
R:Sakurai, Y.; Sullivan, M.; Yamada, Y.
J. Biol. Chem. 261, 6654-6657, 1986
A:Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes
A:Reference number: S19094; MUID:86160699
A:Accession: S19094
A:Molecule type: DNA
A:Residues: 1110-1135;1189-1316;1342-1383;1418-1487 <SAK>
A:Cross-references: EMBL:M13027
R:Schuppan, D.; Timpl, R.; Glanville, R.W.
FEBS Lett. 115, 297-300, 1980
A:Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (ty
A:Reference number: S16909; MUID:80246483
A:Accession: S16909
A:Molecule type: protein

A:Residues: 940-946,'G',948-949,'G',951-955,'G',957;1213-1228,'X',1230-1234,'P',1236-
R:Schuppan, D.; Glanville, R.W.; Timpl, R.
Eur. J. Biochem. 123, 505-512, 1982
A:Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial a
A:Reference number: A25991; MUID:82186723
A:Accession: A25991
A:Molecule type: protein
A:Residues: 940-946,'X',948-949,'X',951-955,'X',957-964,'X',966-991,'X',993-1003,'X',
61,'X',1063-1065,'X',1067-1080,'X',1082-1083,'X',1085-1106,'X',1108-1115,'DE',1118-11
A:Accession: B25991
A:Molecule type: protein
A:Residues: 1173-1181,'X',1183-1184,'X',1186-1187,'X',1189-1205,'Q',1207,'XE',1210-12
3,'SP',1266,'IT',1269,'SK',1272,'DM',1275,'L',1277-1282;1316-1318,'X',1320-1327,'X',1
R:Reber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.
Eur. J. Biochem. 139, 401-410, 1984
A:Title: Subunit structure and assembly of the globular domain of basement-membrane c
A:Reference number: S17801; MUID:84132058
A:Accession: S17801
A:Molecule type: protein
A:Residues: 1435-1443 <MEB>
C:Genetics: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3
A:Note: the list of introns may be incomplete
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: basement membrane; cell binding; coiled coil; duplication; extracellular
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-1669/Product: collagen alpha 1(IV) chain #status predicted <NAT>
F:28-162/Domain: 75 <7SD>
F:163-1440/Domain: collagenous, triple helix <COL>
F:597-599/Region: cell attachment (R-G-D) motif
F:781-783/Region: cell attachment (R-G-D) motif
F:917-919/Region: cell attachment (R-G-D) motif
F:968-970/Region: cell attachment (R-G-D) motif
F:1441-1669/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F:1553-1669/Region: duplication
F:31,36,39,41,434,467,470/Disulfide bonds: Interchain #status predicted
F:126/Binding site: carboxylate (Asn) (covalent) #status predicted
F:971,974,977,986,989,1001,1007,1019,1022,1031,1037,1040,1055,1060,1063,1075,1078,109
92,1298,1310,1313,1322,1337,1346,1349,1422,1431,1437,1440/Modified site: hydroxy
F:1214,1424/Modified site: 4-hydroxyproline (Pro) #status experimental
F:1304/Modified site: 5-hydroxylysine (Lys) #status experimental
F:1505-1511,1616-1622/Disulfide bonds: #status predicted

Query Match	80.8%	Score 315	DB 1	Length 1669
Best Local Similarity	76.1%	Pred. No. 2,7e-29		
Matches 54	Conservative 10	Mismatches 7	Indels 0	Gaps 0

QY 1 ORAHGDLGTLGSCLOEFTMPFLFCNVNVCNPNASNDYSYMLSTPALPMNAPITGR 60
DB 1480 ERHAGDGLTAGSCLERFSTMPFLFCNINVCNPNASNDYSYMLSTPALPMNAPITGR 1539

QY 61 ALPEYTSRCV 71
DB 1540 NRPFTSRCAV 1550

RESULT 8
A55267
collagen alpha 5(IV) chain - dog (fragment)
C:Species: Canis lupus familiaris (dog)
C:Date: 06-Feb-1995 #sequence_revision 06-Feb-1995 #text_change 13-Aug-1999
C:Accession: A55267
R:Zheng, K.; Thorner, P.S.; Mariano, P.; Bauml, R.; McInnes, R.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 3989-3993, 1994
A:Title: Canine X chromosome-linked hereditary nephritis: a genetic model for human X
en type IV.
A:Reference number: A55267; MUID:94224868
A:Accession: A55267
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-754 <ZHE>

RESULT 11
B61228
collagen alpha 1(IV) chain - rabbit (fragment)
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 17-Mar-1999
C:Accession: B61228
R.Yamaguchi, N.; Sato, N.; Ko, J.S.; Ninomiya, Y.
Invest. Ophthalmol. Vis. Sci. 32, 2924-2930, 1991
A:Title: Cloning of alpha1(IV) and alpha2(IV) collagen cDNAs from rabbit corneal endothelium
A:Reference number: A61228; MUID:92010685
A:Accession: B61228
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-258 <YAM>
C:Superfamily: collagen alpha 1(IV) chain

R:Messel, G.M, Etkin, M.; Benson, S.
Dev. Biol. 148, 261-272, 1991
A:Title: Primary mesenchyme cells of the sea urchin embryo require an autonomously pr
A:Reference number: A43903; MUID:92038439
A:Accession: A43903
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 'P', 633<1537, 'G' <WES>
A:Cross-references: GB:564572; NID:9238616; PIDN:AAB20270.1; PID:9238617
A>Note: sequence extracted from NCBI backbone (NCBIN:64572, NCBI:P:64573)
R:Venkatesan, M.; De Pablo, F.; Vogel, G.; Simpson, R.T.
Proc. Natl. Acad. Sci. U.S.A. 83, 3351-3355, 1986
A:Title: Structure and developmentally regulated expression of a Strongylocentrotus f
A:Reference number: A23940; MUID:86205894
A:Accession: A23940
A:Molecule type: DNA
A:Residues: 742-812 <VEN>
A:Cross-references: EMBL:M13206
C:Superfamily: collagen alpha 1(IV) chain
C:Keywords: collod coll, extracellular matrix; glycoprotein; trimer; triple helix
F:29-161/Domain: amino-terminal nonhelical, 7S <750>
F:162-1523/Region: interrupted helical
F:1524-1752/Domain: carboxyl-terminal nonhelical, NC1 <NC1>
F:1524-1634/Domain: collagen IV carboxyl-terminal repeat <CT1>
F:1644-1748/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:123/Modified site: allylsine (lys) #status predicted

RESULT 13

hypothetical protein F01G12.5a - *Caenorhabditis elegans*
C;Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C/Accession: T29350
R:Wu, X.; Le, T.T.
Submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid F01G12.
A:Reference number: 220611
A:Accession: T29350
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1758 <WUX>
A:Cross-references: EMBL:U53342; PIDN:AAA96216.1; GSPDB:GN00028; CESP:F01G12.5a
A:Experimental source: strain Bristol N2; clone F01G12
A:Genetics:
A:Gene: CESP:F01G12.5a
A:Map position: X
A:Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 264/3; 303/3; 358/3; 449/2; 736/3
C:Superfamily: collagen alpha 1(IV) chain

Very Match 65.9%; Score 257; DB 2; Length 1758;
Best Local Similarity 64.8%; Pred. No. 2.3e-22;
Matches 46; Conservative 10; Mismatches 13; Indels 2; Gaps 1;
OY 1 ORAHGODLGTGSCLOFRTTPEFCNVNDVCNFSARNDSYWLSTPALMPNMAPITGR 60
DB 1566 EKSHNODLGHAGSCLOFRTTPEFCNVNDVCNFSARNDSYWLSTSEAIIP--MMPVNER 1623
OY 61 ALEPYISRCTV 71
DB 1624 EIEPYISRCAY 1634

RESULT 14
T29351
collagen alpha 2(IV) chain precursor let-2 - Caenorhabditis elegans
N:Alternate names: collagen alpha 2(IV) chain precursor clb-1
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C/Accession: T29351
R:Wu, X.; Le, T.T.
Submitted to the EMBL Data Library, April 1996
A:Description: The sequence of C. elegans cosmid F01G12.
A:Reference number: 220611
A:Accession: T29351
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1759 <WUX>
A:Cross-references: EMBL:U53342; PIDN:AAA96215.1; GSPDB:GN00028; CESP:F01G12.5a
A:Experimental source: strain Bristol N2; clone F01G12
A:Genetics:
A:Gene: CESP:F01G12.5a
A:Map position: X
A:Introns: 8/2; 26/3; 47/3; 81/1; 144/1; 202/3; 228/3; 265/3; 304/3; 359/3; 450/2; 737/3
C:Superfamily: collagen alpha 1(IV) chain

Query Match 65.9%; Score 257; DB 2; Length 1759;
Best Local Similarity 64.8%; Pred. No. 2.3e-22;
Matches 46; Conservative 10; Mismatches 13; Indels 2; Gaps 1;

OY 1 ORAHGODLGTGSCLOFRTTPEFCNVNDVCNFSARNDSYWLSTPALMPNMAPITGR 60
DB 1567 EKSHNODLGHAGSCLOFRTTPEFCNVNDVCNFSARNDSYWLSTSEAIIP--MMPVNER 1624
OY 61 ALEPYISRCTV 71
DB 1625 EIEPYISRCAY 1635

RESULT 15
S16366
collagen alpha 2(IV) chain precursor - pig roundworm
C:Species: Ascaris suum (pig roundworm)

C>Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C/Accession: S16366
R:Petitt, J.; Kingston, I.B.
J. Biol. Chem. 266, 16149-16156, 1991
A:Title: The complete primary structure of a nematode alpha-2(IV) collagen and the pa
A:Reference number: S16366; MUID:91340768
A:Accession: S16366
A:Molecule type: mRNA
A:Residues: 1-1763 <JIB>
A:Cross-references: GB:M67507; NID:g159648; PIDN:AAA18014.1; PID:g159649
C:Genetics:
A:Introns: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1; 963/
A:Superfamily: collagen alpha 1(IV) chain
C:Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disul
F:1-26/Domain: signal sequence #status predicted <Sig>
F:27-1763/Product: collagen alpha 2(IV) chain #status predicted <MAT>
F:27-42/Domain: non-collagenous NH1 #status predicted <NH1>
F:43-1529/Domain: collagenous #status predicted <COL>
F:197-199/Region: cell attachment (R-G-D) motif
F:1530-1763/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>
F:1530-1638/Domain: repeat NC1 #status predicted <NC11>
F:1639-1763/Domain: repeat NC1 #status predicted <NC12>
F:31,34,39,41,536,539/Disulfide bonds: interchain #status predicted
F:126/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:1593-1599,1702-1709/Disulfide bonds: #status predicted

Query Match 65.6%; Score 256; DB 2; Length 1763;
Best Local Similarity 64.8%; Pred. No. 3e-22;
Matches 46; Conservative 11; Mismatches 12; Indels 2; Gaps 1;
OY 1 ORAHGODLGTGSCLOFRTTPEFCNVNDVCNFSARNDSYWLSTPALMPNMAPITGR 60
DB 1568 EKSHNODLGHAGSCLOFRTTPEFCNVNDVCNFSARNDSYWLSTTA--PIPMVPVSEG 1625
OY 61 ALEPYISRCTV 71
DB 1626 EIEPYISRCAY 1636

Search completed: April 13, 2001, 15:42:41
Job time: 222 sec

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 14, 2001, 08:56:23 ; Search time 13.78 Seconds
(without alignments)
94.756 Million cell updates/sec

Title: US-09-543-371-10_COPY_185_203
Perfect score: 115
Sequence: 1 SPFLCHGRGRCNYNSY 19

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: 1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	115	100.0	1670	1 CGH03B	collagen alpha 3(I)
2	113	98.3	161	2 S49488	collagen alpha 3(I)
3	113	98.3	246	2 I48302	collagen alpha 3(I)
4	113	98.3	471	2 A39024	collagen alpha 3(I)
5	109	94.8	79	2 C43928	probable collagen
6	107	93.0	253	2 I48304	collagen alpha 5(I)
7	107	93.0	754	2 A55267	collagen alpha 5(I)
8	107	93.0	1691	1 S22917	collagen alpha 5(I)
9	106	92.2	58	2 B43928	probable collagen
10	104	90.4	258	2 B61228	collagen alpha 1(I)
11	104	90.4	1669	1 CGH04B	collagen alpha 1(I)
12	104	90.4	1669	1 CGH04B	collagen alpha 1(I)
13	92	80.0	1744	2 S40991	collagen alpha 1(I)
14	88	76.5	1752	2 A45407	collagen alpha 1(I)
15	81	70.4	1747	2 A54121	collagen alpha 4(I)
16	80.5	70.0	775	2 A61228	collagen alpha 2(I)
17	80.5	70.0	1712	1 CGH02B	collagen alpha 2(I)
18	76.5	66.5	1707	2 A33526	collagen alpha 2(I)
19	74.5	64.8	1691	1 CGH06B	collagen alpha 6(I)
20	73.5	63.9	453	2 S18804	collagen alpha 4(I)
21	73.5	63.9	623	2 A45137	collagen alpha 4(I)
22	73.5	63.9	1690	1 CGH01B	collagen alpha 4(I)
23	72.5	63.0	1763	2 S16366	collagen alpha 4(I)
24	70.5	61.3	312	2 I48303	collagen alpha 4(I)
25	70	60.9	1761	2 T13990	collagen type IV
26	68.5	59.6	261	2 A34476	collagen alpha 2(I)
27	68.5	59.6	1758	2 T29351	hypothetical prote
28	68.5	59.6	1759	2 T29351	collagen alpha 2(I)
29	59.5	51.7	1775	2 A31893	collagen alpha 1(I)

30	56	48.7	903	2 S60257	meltrin alpha - mo
31	49.5	43.0	985	2 S67255	REVI protein - yea
32	48	41.7	176	2 B70606	hypothetical prote
33	48	41.7	660	2 S71949	metalloproteinase
34	47	40.9	204	2 T08072	proteinase inhibit
35	47	40.9	600	2 I49281	fertilin alpha pre
36	47	40.9	825	2 S55060	fertilin alpha-ii
37	47	40.9	905	2 S55059	fertilin alpha-i -
38	46	40.0	240	2 T14978	hypothetical prote
39	46	40.0	703	2 G82249	asma protein VC103
40	46	40.0	740	2 A71141	hypothetical prote
41	46	40.0	777	2 I48100	ADAM 5 protein pre
42	46	40.0	1203	2 A49175	cell-fate determin
43	46	40.0	2471	2 A49128	hypothetical prote
44	45	39.1	230	2 T31722	hypothetical prote
45	45	39.1	378	2 T18486	hypothetical prote

ALIGNMENTS

RESULT 1
CGH03B
collagen alpha 3(IV) chain precursor, long splice form - human
N: Alternate names: Goodpasture antigen; procollagen alpha 3(IV) chain long splice for
C: Species: Homo sapiens (man)
C: Date: 28-Oct-1994 #sequence, revision 03-Oct-1995 #text, change 22-Jun-1999
C: Accession: A54763, A43928, A44043, A45971, A39286
R: Mariyama, M.; Leinonen, A.; Mochizuki, T.; Tyrgvason, K.; Reders, S.T.
J. Biol. Chem. 269, 23013-23017, 1994
A: Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpress
A: Reference number: A54763; MID:94364994
A: Accession: A54763
A: Molecule type: mRNA
A: Residues: 1-1670 <MAR>
A: Cross-references: GB:X80031; NID:9577563; PID:9577564
A: Experimental source: Kidney
R: Turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
J. Clin. Invest. 89, 592-601, 1992
A: Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the
A: Reference number: A43928; MID:92147878
A: Accession: A43928
A: Molecule type: mRNA
A: Residues: 1331-1524, 1526-1670 <TUR>
A: Cross-references: GB:M81379
A: Experimental source: Kidney
R: Quinones, S.; Bernat, D.; Garcia-Segor, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 267, 19780-19784, 1992
A: Title: Exon/intron structure of the human alpha-3(IV)-gene-encompassing the Goodpas
A: Reference number: A44043; MID:93015826
A: Accession: A44043
A: Molecule type: DNA; mRNA
A: Residues: 1386-1670 <CUR>
A: Cross-references: GB:M92993; NID:9177895; PID:AA21610.1; PID:9177896
A: Note: sequence extracted from NCBI backbone (NCBIP:115592)
R: Quinones, S.; Bernat, D.; Garcia-Segor, M.; Elena, S.F.; Saus, J.
J. Biol. Chem. 269, 17258, 1994
A: Reference number: A44738; MID:94274734
A: Contents: annotation; erratum; correction to intronic sequence in A44043
R: Bernat, D.; Quinones, S.; Saus, J.
J. Biol. Chem. 268, 12090-12094, 1993
A: Title: The human alpha-3(IV) chain is alternatively spliced.
A: Reference number: A45971; MID:93280184
A: Accession: A45971
A: Status: nucleic acid sequence not shown
A: Molecule type: mRNA
A: Residues: 1427-1444 <BER>
A: Note: sequence extracted from NCBI backbone (NCBIP:133363); sequence incorrectly id
R: Morrison, K.E.; Mariyama, M.; Yang-Feng, T.L.; Reders, S.T.
Am. J. Hum. Genet. 49, 545-554, 1991
A: Title: Sequence and localization of a partial cDNA encoding the human alpha3 chain
A: Reference number: A39786; MID:915353570

```

A:Accession: A39786
A:Molecule type: mRNA
A:Residues: 1453-1593, 'A', 1595-1670 <MOR>
A:Cross-references: GB:S55790; NID:g224418; PIDN:AA19637.1; PID:g234419
A:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
ed and subsequently O-glycosylated.
C:Comment: In Goodpasture's syndrome, an autoimmune response develops against an epitope
C:Genetics:
A:Gene: GDB:COL4A3
A:Cross-references: GDB:128351; OMIM:120070
A:Map position: 2q36-2q37
A:Introns: 1385/1; 1418/1; 1488/1; 1547/2; 1585/3; 1643/2 #status incomplete
A:Note: the alpha 3(IV) and alpha 4(IV) chain genes are encoded on opposite strands with
C:Complex: This minor type IV collagen is thought to form a heterotrimer of two alpha 3
among trimer amino-terminal domains (with disulfide and desmosine cross-links), dimeric
er associations in the interrupted helical domain (with disulfide and desmosine cross-
C:Function:
A:Description: minor structural component of extracellular basement membrane in kidney
C:Superfamily: collagen alpha 1(IV) chain
A:Keywords: alternative splicing; status predicted <SIG>
A:Domain: signal sequence #status predicted <SIG>
F:29-42/Domain: amino-terminal nonhelical, NHL <NHL>
F:43-1438/Region: interrupted helical
F:791-793/Region: cell attachment (R-G-D) motif
F:996-998/Region: cell attachment (R-G-D) motif
F:1154-1156/Region: cell attachment (R-G-D) motif
F:1306-1308/Region: cell attachment (R-G-D) motif
F:1345-1347/Region: cell attachment (R-G-D) motif
F:1432-1434/Region: cell attachment (R-G-D) motif
F:1438-1670/Domain: carboxyl-terminal nonhelical, NCL <NCL>
F:1451-1551/Domain: collagen IV carboxyl-terminal repeat <CT1>
F:1561-1665/Domain: collagen IV carboxyl-terminal repeat <CT2>
F:151,33,39,41,135,422,476,682,722,809,1387/Disulfide bonds: interchain #status predi
F:253/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:1460-1548,1493-1551/Disulfide bonds: (or 1460-1551, 1493-1548) #status predicted
F:1505-1511,1616-1622/Disulfide bonds: #status predicted
F:1570-1662,1604-1665/Disulfide bonds: (or 1570-1665, 1604-1662) #status predicted

Query Match 100.0%; Score 115; DB 1; Length 1670;
Best Local Similarity 100.0%; Pred. No. 9,1e-09;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SPFLECHRGTCNYNSY 19
Db 1611 SPFLECHRGTCNYNSY 1629

R: 2
S49488
collagen alpha 3(IV) chain - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 13-Aug-1999
C:Accession: S49488
R:Oberbauer, I.
submitted to the EMBL Data Library, October 1994
A:Description: Cloning of the NCI domains for the minor collagen IV chains of mouse via
els.
A:Reference number: S49487
A:Accession: S49488
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-161 <OEB>
A:Cross-references: EMBL:X82205; NID:g559472; PIDN:CA57689.1; PID:g559916
C:Superfamily: collagen alpha 1(IV) chain

Query Match 98.3%; Score 113; DB 2; Length 161;
Best Local Similarity 94.7%; Pred. No. 2,6e-09;
Matches 18; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SPFLECHRGTCNYNSY 19
Db 187 SPFLECHRGTCNYNSY 205

RESULT 4
A39024
collagen alpha 3(IV) chain - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 13-Aug-1999
C:Accession: A39024; S20672; S17802; A35167; C39419; S13747; S20815
R:Morrisson, K.E.; Germino, G.G.; Reeders, S.T.
J. Biol. Chem. 266, 34-39, 1991.
A:Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding t
A:Reference number: A39024; MUID:91093146
A:Accession: A39024
A:Molecule type: mRNA
A:Residues: 1-471 <MOR>
A:Cross-references: EMBL:M63139; NID:g162886; PIDN:AA62708.1; PID:g162887
R:Butkowski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.
J. Biol. Chem. 262, 7874-7877, 1987
A:Title: Localization of the Goodpasture epitope to a novel chain of basement membran
A:Reference number: S18432; MUID:87222419
A:Accession: S20672
A:Molecule type: protein
A:Residues: 227-228, 'X', 230-244 <BUT>
R:Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Outones, S.; Hudson, B.G.
J. Biol. Chem. 263, 13374-13380, 1988
A:Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collag
A:Reference number: S17802; MUID:88330844
A:Accession: S17802
A:Molecule type: protein
A:Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>
R:Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.
J. Biol. Chem. 265, 5466-5469, 1990
A:Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of t
A:Reference number: A35167; MUID:90202779
A:Accession: A35167
A:Molecule type: protein
A:Residues: 236-258 <GUN>
R:Gunwar, S.; Ballester, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.;
J. Biol. Chem. 266, 15318-15324, 1991
A:Title: Glomerular basement membrane. Identification of dimeric subunits of the nonc
A:Reference number: A39419; MUID:91332055
A:Accession: C39419
A:Molecule type: protein
A:Residues: 236-255 <G02>

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2001, 15:44:43 ; Search time 13.93 Seconds
(without alignments)
550.275 Million cell updates/sec

Title: US-09-543-371-10_copy_54_124
Perfect score: 390
Sequence: 1 QRAHGQDLGTLGSLQGRFTT.....MNNAPITGRALPEYISRCTV 71

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

number of hits satisfying chosen parameters: 17386

Minimum DB seq length: 0
Maximum DB seq length: 71

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	78.5	20.1	58	B43928	probable collagen
2	45.5	11.7	61	D70246	hypothetical prote
3	45	11.5	61	TIEH	proteinase inhibit
4	44.5	11.4	66	G82626	hypothetical prote
5	43.5	11.2	52	H72124	hypothetical prote
6	43	11.0	30	D81561	hypothetical prote
7	43	11.0	34	C70227	hypothetical prote
8	43	11.0	70	T13341	hypothetical prote
9	42	10.8	64	C64010	hypothetical prote
10	42	10.8	67	F72226	hypothetical prote
11	41.5	10.6	55	J01540	hypothetical prote
12	41.5	10.6	57	B46554	hypothetical 6.2K
13	41	10.5	29	T17076	growth modulatory
14	41	10.5	60	G82472	NADH dehydrogenase
15	40	10.3	60	A36989	hypothetical prote
16	40	10.3	64	H69358	calicudine - eas
17	39.5	10.1	64	S20961	keratinocyte growt
18	39.5	10.1	69	A64527	hypothetical prote
19	39	10.0	42	S36250	hypothetical prote
20	39	10.0	49	S50999	superoxide dismuta
21	39	10.0	55	B35732	calin-like prote
22	39	10.0	60	S53364	mucin 5AC (clone J
23	39	10.0	70	PL0250	Ig heavy chain V r
24	38.5	9.9	52	W8WLB4	E8 protein - bovin
25	38.5	9.9	55	A42824	triacetylglucosyl
26	38.5	9.9	63	T26855	hypothetical prote
27	38.5	9.9	68	J00034	long neurotoxin ho
28	38	9.7	29	B43937	endo-1,4-beta-xyla
29	38	9.7	2	JT0514	Ig heavy chain V-I

30	38	9.7	45	2	F64801	hypothetical prote
31	38	9.7	50	2	E82816	hypothetical prote
32	38	9.7	58	1	Z5BP83	gene 58 protein -
33	38	9.7	69	2	I49617	gamma-F-crystallin
34	38	9.7	69	4	S58502	probable IAI15 pse
35	37.5	9.6	42	2	A34396	pyrBI leader pepit
36	37.5	9.6	54	2	T37102	probable sigma fac
37	37.5	9.6	54	2	F31436	ovomucoid, third d
38	37.5	9.6	57	2	C46654	growth modulatory
39	37.5	9.6	61	2	D69193	hypothetical prote
40	37.5	9.6	69	2	C71980	hypothetical prote
41	37	9.5	27	2	JC1081	brain natriuretic
42	37	9.5	69	2	A84047	hypothetical prote
43	37	9.5	70	2	E33172	C-ORF-D protein -
44	36.5	9.4	42	2	D82376	hypothetical prote
45	36.5	9.4	65	1	NTSREB	neurotoxin XI - sc

ALIGNMENTS

RESULT 1
B43928
Probable collagen alpha 5(IV) chain - sheep (fragment)
C/Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C/Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-1999
C/Accession: B43928
R/turner, N.; Mason, P.J.; Brown, R.; Fox, M.; Povey, S.; Rees, A.; Pusey, C.D.
J. Clin. Invest. 89, 592-601, 1992
A/Title: Molecular cloning of the human Goodpasture antigen demonstrates it to be the
A/Reference number: A43928; MID:92147878
A/Accession: B43928
A/Status: preliminary; not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-58 <TUR>
C/Superfamily: collagen alpha 1(IV) chain
C/Keywords: collid coll; extracellular matrix; glycoprotein; trimer; triple helix

Query Match 20.18; Score 78.5; DB 2; Length 58;
Best Local Similarity 35.2%; Pred. No. 0.013;
Matches 19; Conservative 8; Mismatches 22; Indels 5; Gaps 2;

QY 22 PFLECHNNDVNCNFASTRDYSWLTSPALMPNNAP-----ITGRALPEYISRCTV 71
Db 3 PFLECHERGTCNYA-NSYSFWLATVDVSDFKPSQSETLQAGDLRTIRISQCV 55

RESULT 2
D70246
hypothetical protein BB11 - Lyme disease spirochete plasmid J/1p38
C/Species: Borrelia burgdorferi (Lyme disease spirochete)
C/Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C/Accession: D70246
R/Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh
son, D.; Peterson, J.; Kerkavag, A.R.; Quackebush, J.; Salzberg, S.; Hanson, M.; Vu
Nature 390, 580-586, 1997
A/Authors: Smith, H.O.; Venter, J.C.
A/Title: Genomic sequence of a Lyme disease spirochete.
A/Reference number: A70100; MID:98065943
A/Accession: D70246
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-61 <KLE>
A/Cross-references: GB:AE000787; NID:g2690175; PIDN:AAC66098.1; PID:g2690190; TTGR:BB
A/Experimental source: strain B31
C/Genetics:
A/Genome: plasmid

Query Match 11.7%; Score 45.5; DB 2; Length 61;
Best Local Similarity 29.8%; Pred. No. 1.2e+02;

	Matches	17;	Conservative	8;	Mismatches	17;	Indels	15;	Gaps	4;
QY	25	FCNV	---NDVCFASRN-----	DYSWLS	TPA--LMP--	MNNA	PITGRAL	EPT	66	
		:	:::			:	:			
Db	5	FCNLCITKNIHIGSLR	RYTFNVYDVAWLS	IKNEDYPPDAK	DIYC	IKGKK	PYI	61		

```

RESULT      3
TIEEH
proteinase inhibitor homolog - European eel
C:Species: Anguilla anguilla (European eel)
C:Date: 30-Jun-1991 #sequence-revision 30-Jun-1991 #text-change 16-Jul-1999
C:Accession: S00630
R:Conlon, J.M.; Thim, L.
Eur. J. Biochem. 174, 149-153, 1988
A:Title: A peptide from the eel pancreas with structural similarity to human pancreatic
A:Reference number: S00630; MUID:8825096
A:Accession: S00630
A:Molecule type: protein
A:Indices: 1-61 <CON>
C:Family: pancreatic secretory trypsin inhibitor; Kazal proteinase inhibitor homol
C:Keywords: pancreas
F:10-61/Domian: Kazal proteinase inhibitor homology <KPI>
F:12-43; 21-40; 29-61/Disulfide bonds: #status predicted

```

Query Match	11.5%;	Score 45;	DB 1;	Length 61;
Best Local Similarity	38.1%;	Pred. No. 1.3e+02;		
Matches	8;	Conservative	4;	Mismatches 9;
				Indels 0;
				Gaps 0;

```
QY      51 PMNMAPITGRALEPYISRCTV 71
      ||| ||: | :|::
Db      22 PMNEAPVCCTDGNITYPNECSL 42
```

```

result      4
682626      Hypothetical protein XF1879 [imported] - xylella fastidiosa (strain 9a5c)
C:Species:  xylella fastidiosa
C:Date:     18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: G82626
R:Anonymus, The xylella fastidiosa Consortium of the Organization for Nucleotide Sequence
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen xylella fastidiosa.

```

A>Note: for a complete list of authors see reference number A59328 below
A:Accession: G82626
A:Genus: preliminary
A:Genule type: DNA
A:Accession: 1-66 <STM>
A:Cross-references: GB:AE004008; GB:AE003849; NID:99106961; PIDN:AAE84685.1; GSPDB:GN001
A:Experimental source: strain 9a6c
A:Simpton, A.J.G.; Relnach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; B
B:Rioson, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carer, I
as Neto, E.; Docena, C.; El-Dorri, H.; Fachinani, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kurmae, E.E.; Laiga
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, H
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; M
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasabi
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.;
A:Reference number: A59328
A:Contents: annotation
A:Contents:
C:Gene: XF1879

Query Match	11.48;	Score 44.5;	DB 2;	Length 66;
Best Local Similarity	32.7%;	Pred. No. 1.7e+02;		
Matches 10; Conservative	6;	Mismatches 18;	Indels 9;	Gaps 3

```
Oy      24 LFCNVNDVCFASFASRNDYSYWLSTPALPMNNAPIT---GRALEPYISRC   69
        |:| | | | | | | | | | | | | | | | | | | | | | | |
Db      8 LYCTQYDV--EARTCSQAAMWBPSSLP----PISEYEDVRILLPHIIVMC   50
```

RESULT 5
 H72124
 Hypoetical protein CP0711 [imported] - Chlamydophil pneumoniae (strains CML029 and C:Species: Chlamydoiphila pneumoniae, Chlamydia pneumoniae
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
 C:Accession: H72124; E81546
 C:Katman, S.; Mitchell, W.; Marathe, R.; Lammell, C.; Fan, J.; Olinger, L.; Grimwood, Nature Genet. 21, 385-389, 1999
 A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
 A:Reference number: A72000; MUID:99206606
 A:Accession: H72124
 A:Molecule type: DNA
 A:Residues: 1-52 <ARN>
 A:Cross-references: GB:AE001591; GB:AE001363; NID:94376311; PIDN:AD18216.1; PID:94377
 A:Experimental source: strain CML029
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
 A:Reference number: A81500; MUID:20150255
 A:Accession: E81546
 A:Molecule type: DNA
 A:Residues: 1-52 <REA>
 A:Cross-references: GB:AE002230; GB:AE002161; NID:g7189624; PIDN:AAF38519.1; PID:g7718
 A:Experimental source: strain AR39, HL cells
 C:Genetics:
 A:Gene: CP00063; CP0711

Query Match	11.2%	Score 43.5;	DB 2;	Length 52;
Best Local Similarity	29.5%;	Pred. No. 1.7e+02;		
Matches	13;	Conservative	6;	Mismatches 20;
				Indels 5;
				Gaps 2;

OY 26 CAVNDYCNFASRNDYSYLSTPALMNMNAPITGRALPEYISRC 69
| | : : | : | : | : | :
Db 5 CHNNCLY---DSRRHSPPGL-PLTFPPPYFTLGIETGRC 43

RESULT 6
D81561
hypothetical protein CP0591 [imported] - Chlamydomonas pneumoniae (strain AR39)
C:Species: Chlamydomonas pneumoniae; Chlamydomonas pneumoniae
C:Date: 31-Mar-2000 #sequence-revision 31-Mar-2000 #text-change 11-May-2000
C:Accession: D81561
R:Read: T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberger, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MOpn and Chlamydia pneumoniae AR39
A:Reference number: AB1500; MUID:20150255
A:Accession: D81561
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-30 <RBA>
A:Cross-references: GB:AE002217; GB:AE002161; NID:g7189493; PIDN:AAF38409.1; PID:g718
A:Experimental source: strain AR39, HL cells
C:Genetics:
A:Gene: CP0591

Query Match	11.0%;	Score 43;	DB 2;	Length 30;
Best Local Similarity	50.0%;	Pred. No. 1.1e+02;		
Matches	10; conservative	1; Mismatches	9; indels	0; Gaps 0;

```
QY      5 GQDLGTLGSCLOQRTTMPFL 24
          || | ||| : | |
Db      6 GQARVTAGSCLAKQETETLL 25
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 13, 2001, 15:47:13 ; Search time 14.26 Seconds
(without alignments)
25.597 Million cell updates/sec

Title: US-09-543-371-10.COPY_185_203

Perfect score: 115

Sequence: 1 SPFLCHGRGTCMYNSY 19

Scoring table: BLOSUM62

Searched: 185757 seqs, 19210857 residues

number of hits satisfying chosen parameters: 95100

Minimum DB seq length: 0

Maximum DB seq length: 19

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCMYUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	28.7	15	1	US-08-267-092A-57
2	33	28.7	15	1	US-08-267-092A-64
3	33	28.7	15	2	US-08-540-412-58
4	33	28.7	15	2	US-08-540-412-59
5	33	28.7	15	2	US-08-540-412-66
6	33	28.7	15	4	US-09-051-342-58
7	33	28.7	15	4	US-09-051-342-59
8	33	28.7	15	4	US-09-051-342-59
9	33	28.7	15	4	US-08-468-161-58
10	33	28.7	15	4	US-08-468-161-59
11	33	28.7	15	4	US-08-468-161-66
12	33	28.7	15	4	US-09-051-759-58
13	33	28.7	15	4	US-09-051-759-59
14	33	28.7	15	4	US-09-051-759-66
15	33	28.7	15	5	US-09-051-759-66
16	33	28.7	15	5	US-09-051-759-66
17	33	28.7	15	5	US-09-051-759-66
18	33	28.7	15	5	US-09-051-759-66
19	33	28.7	15	5	US-09-051-759-66
20	33	28.7	15	5	US-09-051-759-66
21	33	28.7	15	5	US-09-051-759-66
22	33	28.7	15	5	US-09-051-759-66
23	33	28.7	15	5	US-09-051-759-66
24	33	28.7	15	5	US-09-051-759-66
25	33	28.7	15	5	US-09-051-759-66
26	33	28.7	15	5	US-09-051-759-66
27	33	28.7	15	5	US-09-051-759-66
28	33	28.7	15	5	US-09-051-759-66
29	33	28.7	15	5	US-09-051-759-66
30	33	28.7	15	5	US-09-051-759-66

28	30	26.1	12	3	US-08-467-580-125	Sequence 125, App
29	30	26.1	12	5	US-08-467-580-125	Sequence 125, App
30	30	26.1	15	1	US-08-267-092A-7	Sequence 2, Appl
31	30	26.1	15	1	US-08-267-092A-58	Sequence 58, Appl
32	30	26.1	15	1	US-08-267-092A-65	Sequence 65, Appl
33	30	26.1	15	2	US-08-540-412-46	Sequence 46, Appl
34	30	26.1	15	2	US-08-540-412-60	Sequence 60, Appl
35	30	26.1	15	2	US-08-540-412-65	Sequence 65, Appl
36	30	26.1	15	2	US-08-540-412-65	Sequence 65, Appl
37	30	26.1	15	4	US-09-051-342-2	Sequence 2, Appl
38	30	26.1	15	4	US-09-051-342-46	Sequence 46, Appl
39	30	26.1	15	4	US-09-051-342-60	Sequence 60, Appl
40	30	26.1	15	4	US-09-051-342-65	Sequence 65, Appl
41	30	26.1	15	4	US-08-468-161-2	Sequence 2, Appl
42	30	26.1	15	4	US-08-468-161-46	Sequence 46, Appl
43	30	26.1	15	4	US-08-468-161-60	Sequence 60, Appl
44	30	26.1	15	4	US-08-468-161-65	Sequence 65, Appl
45	30	26.1	15	4	US-09-051-759-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-267-092A-57
Sequence 57, Application US/08267092A
Patent No. 559686
GENERAL INFORMATION:
APPLICANT: Defeo-Jones, Deborah
APPLICANT: Garsky, Victor M.
APPLICANT: Jones, Raymond E.
APPLICANT: Oliff, Allen I.
TITLE OF INVENTION: NOVEL PEPTIDES
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESSES:
ADDRESSEE: David A. Muthard
STREET: 126 E. Lincoln Avenue, P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065-0900
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/267, 092A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Muthard, David A.
REGISTRATION NUMBER: 35,297
REFERENCE/DOCKET NUMBER: 19253
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3903
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
US-08-267-092A-57
Query Match 28.7%; Score 33; DB 1; Length 15;
Best Local Similarity 54.5%; Pred. No. 79;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 8 GRGTCNYSNS 18
| | | : | | | :
Db 1 GKGISQYSNS 11

RESULT 2
US-08-267-092A-64
Sequence 64, Application US/08267092A
Patent No. 559686
GENERAL INFORMATION:
APPLICANT: Defeo-Jones, Deborah
APPLICANT: Garsky, Victor M.
APPLICANT: Jones, Raymond E.
APPLICANT: Oliff, Allen I.
TITLE OF INVENTION: NOVEL PEPTIDES
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: David A. Muthard
STREET: 126 E. Lincoln Avenue, P.O. Box 2000
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065-0900
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/267,092A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Muthard, David A.
REGISTRATION NUMBER: 35,297
REFERENCE/DOCKET NUMBER: 19253
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3903
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
US-08-267-092A-64

Query Match 28.7%; Score 33; DB 1; Length 15;
Best Local Similarity 54.5%; Pred. No. 79;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
OY 8 GRGTCNYSNS 18
| | | : | | | :
Db 1 GKGISQYSNS 11

RESULT 3
US-08-540-412-58
Sequence 58, Application US/08540412
Patent No. 586679
GENERAL INFORMATION:
APPLICANT: Defeo-Jones, Deborah
APPLICANT: Feng, Dong-Mei
APPLICANT: Garsky, Victor M.
APPLICANT: Jones, Raymond E.
APPLICANT: Oliff, Allen I.
TITLE OF INVENTION: NOVEL PEPTIDES

NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID A. MUTHARD
STREET: 126 E. Lincoln Avenue, P.O. BOX 2000
CITY: RAHWAY
STATE: NEW JERSEY
COUNTRY: U.S.A.
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/540,412
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Muthard, David A.
REGISTRATION NUMBER: 35,297
REFERENCE/DOCKET NUMBER: 192531C
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3903
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 15 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
US-08-540-412-58

Query Match 28.7%; Score 33; DB 2; Length 15;
Best Local Similarity 54.5%; Pred. No. 79;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
OY 8 GRGTCNYSNS 18
| | | : | | | :
Db 1 GKGISQYSNS 11

RESULT 4
US-08-540-412-59
Sequence 59, Application US/08540412
Patent No. 586679
GENERAL INFORMATION:
APPLICANT: Defeo-Jones, Deborah
APPLICANT: Feng, Dong-Mei
APPLICANT: Garsky, Victor M.
APPLICANT: Jones, Raymond E.
APPLICANT: Oliff, Allen I.
TITLE OF INVENTION: NOVEL PEPTIDES
NUMBER OF SEQUENCES: 194
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID A. MUTHARD
STREET: 126 E. Lincoln Avenue, P.O. BOX 2000
CITY: RAHWAY
STATE: NEW JERSEY
COUNTRY: U.S.A.
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/540,412
FILING DATE: